UAD Library 01382 308833
294452

Mechanistic Modelling of Bioavailability: Putting Soil Microbiology in its Pore Scale Context

Jennifer M. Blair

UNIVERSITY of ABERTAY
DUNDEE LIBRARY
11 OCT 2007

Ph.D. 2007

Mechanistic Modelling of Bioavailability: Putting Soil Microbiology in its Pore Scale Context

A thesis submitted in partial fulfilment of the requirements of the University of Abertay Dundee for the degree of Doctor of Philosophy

by

Jennifer M. Blair B.Sc. (Hons) Mathematical Science

University of Abertay Dundee

School of Computing and Creative Technologies

March 2007

I certify that this thesis is the true and accurate version of the thesis approved by the examiners.

[redacted]

(Director of Studies)

Date 7/9/07

Declaration

I herby declare that this thesis has been composed by myself and that it has not been accepted in any previous application for a degree. The work of which it is a record, is my own, unless otherwise stated. All verbatims have been distinguished by quotation marks and sources of information specifically acknowledged by means of references.

Jennifer M. Blair

Acknowledgements

I would like to express my sincere thanks to:

My supervisors Professor John Crawford and Professor Iain Young and the BBSRC for funding my studentship and thus providing me with the chance to undertake a doctoral programme.

Dr Ruth Falconer and Dr XiaoXian Zhang for their continuous advice and support in all matters relating to diffusion. Dr Jim Bown for his coaching on how not to react to uncertainty and circumstance.

My friends in SIMBIOS and CCT for their continued support and humour that has keep me going the past three years.

I would also like to thank my husband, Mum, brother, Gran and in-laws for their encouragement and support as well as keeping me sane during these past 3 years. Finally, I would especially like to thank my Dad for his incessant intellectual and challenging conversations regarding all things soil.

Table of Contents

List of Figures

List of Tables

Abstract

The soil-microbe complex was investigated in terms of the relationship between the physical soil structure and related functional characteristics. The use of CT scanning technology provided a means of obtaining high-resolution digital reconstructions of the internal soil structures in three dimensions thus allowing, for the first time, the internal architecture of soil to be investigated. The investigation of soil structure was facilitated through the development of a theoretical modelling framework that acted as a virtual laboratory. A suite of metrics was implemented within the modelling framework to allow the quantification of the physical soil structure. The analysis of a range of CT imaged soil structures via the quantification metrics showed that the micro-structure of soil can span a range of porosities and textures. The metric values obtained for individual structures characterise the features of the structures that play an influential role in mediating the functional value of a sample.

A model that simulates soil structure in three dimensions from two-dimensional data is shown to reproduce the key structural characteristics of soil at the micro-scale. Data extracted from the CT imaged soil samples were extracted to parameterise the model and thus the simulated soil structures were validated against the original CT imaged structure using the quantification metrics. In general, the model adequately reconstructed the variation found across the CT imaged samples. However, for low porosity samples the complex pore architectures were not captured.

The soil structures were used to simulated oxygen transport through the microenvironment to microbial active sites thus allowing the examination of the impact of physical structure on microbial respiration. The results presented show that soil structure is the determining factor in bulk respiration of the soil microenvironment.

Chapter 1 Introduction and Aims

1.1 The Soil Microenvironment

Soil is fundamental to all life on Earth. However, the processes and interactions that take place within the system, at multiple scales, are still poorly understood. The aim of this thesis is to investigate how the functional characteristics of soil at the micro-scale, such as resident microbial activity, are influenced by the physical structure of soil. Progress requires that we not only recognise the complexity, but also the order that is present in the soil ecosystem at the micro-scale. This is the key to gaining insight into the processes that influence and regulate the biotic activity as the physical structure governs the transport of vital resources around the microenvironment. The theoretical framework of soil presented by Tisdall and Oades (1982) gives a qualitative hierarchical representation of the soil system where physical, chemical and biological interactions all contribute to the complexity of the soil system at the micro-scale. Tisdall and Oades (1982) show that the stability of soil aggregates is influenced by the quantity and distribution of organic matter within the soil. This framework has been developed by many studies both experimental and theoretical (Dorioz et. al., 1993, Feeney, et. al, 2006, Nunan, et. al., 2006).

Laboratory experiments have provided evidence for the role of micro-organisms in changing the functional properties of the soil by changing the physical structure of the soil into more effective or favourable conditions (Chenu, 1993, Feeney et. al, 2006). The functional role of roots, fungi and bacteria in regulating the soil ecosystem at the micro-scale has been investigated and it is shown that each of these contributes to the ecosystem regulation at different spatial scales (Chenu, 1993). Chenu (1993) and Dorioz et. al. (1993) show that extracellular polysaccharides (EPS), which are secreted by microbial cells, can alter the physical structure of soil at the particle level and increase water retention, thus the production of EPS by soil microbes could influence microbial function and survival. Microbial life in soil relies on the physical structure of soil not only in terms of providing a natural habitat and physical protection from predation, but also to provide transport pathways for vital resources

such as water, oxygen and nutrients. The pore networks of a soil microenvironment provide these transport pathways and influence the level of microbial activity that can be sustained. An open, more porous structure facilitates oxygen transport and aids the flow of water through a structure. However, if water becomes limiting, the open structure may become desiccated and predation rates increase.

With the advancement of computed tomography (CT) technology, it is now possible to obtain three-dimensional (3D) images of the interior structure of soil at scales relevant to microbial processes (Feeney et. al. 2006). These images provide, for the first time, data that allows the soil microstructure to be analysed in detail and crucially in three dimensions. Numerous and varied techniques exist that are used to quantify soil structure. These are predominantly metrics that describe the pore structure, such as the porosity, connectivity and various fractal dimensions. The specific metric values can then be used to predict the consequences for the functional characteristics of the structure, such as oxygen availability. These quantification metrics are used a great deal in describing soil functional properties due to the relative difficulty of gaining reliable data from laboratory experiments.

The relation between physical habitat and ecosystem functioning is crucial to understanding how the system as a whole is regulated. The integrated soil ecosystem, i.e. the combination of the physical and biological elements may be viewed as a complex adaptive system (CAS). Levin (1998) defines the study of a complex adaptive system to be "a study of how complicated structures and patterns of interaction can arise from disorder through simple but powerful rules that guide change". The relationships and interactions between the component parts within a CAS are shown to lead to the emergence of large-scale system properties (Crawford, et. al., 2005, Levin, 1998). The relationships and interactions between the biotic and abiotic elements of the soil ecosystem are fundamental to understanding the self-organisation that is thought to occur in the soil microenvironment (Wolf and Holvoet, 2005, Feeney et. al., 2006, Young and Crawford, 2004).

1.2 Problem Definition

In order to investigate the relationship between the physical soil structure and functional characteristics, a theoretical modelling framework is required. Based on the challenges of integrating the physical and biological aspects of the soil microenvironment, three main objectives for the modelling framework emerge. First, the explicit three-dimensional representation of soil structure and microbial distributions is required. The data resulting from CT scanning of soil samples provides a 3D grid of volume elements that indicate the state (i.e. pore or solid) of the structure element. These data are increasingly available allowing the internal structure of soil samples to be analysed at the micro-scale. Data relating to the spatial distribution of microbes is available from 2D biological thin-sections. Although these data are limited to two dimensions, they also provide some structural information and can be used as the basis to develop 3D models of the relative distribution of soil structure and microbes. Therefore, the modelling framework must have the ability to handle 2D structural data also. Using models based on two-dimensional data with input taken from 3D CT data, the models ability to reproduce 3D structure can be tested.

Secondly, in order to examine the relationship between the functional characteristics and physical structural properties, the physical soil structure must be quantified in some manner. As stated above, there are numerous and varied methods of quantifying soil structure (see Section 2.2), mainly in terms of metric values describing some aspect of the structure (i.e. the pore space or solid matrix). Therefore, a range of quantitative tools will be required to fully characterise the soil structure samples under investigation and thus provide the means of relating structure to function.

Thirdly, to enable the relation between physical structure and biological function to be investigated, the modelling framework must possess the ability to simulate some functional element of the soil microenvironment. Previous studies have highlighted the importance of soil aeration for microbial activity (Dorioz et. al, 1993, Young and Crawford, 2001). Therefore, it seems reasonable to assume that oxygen availability within a structure will be a key resource that will influence the nature of resident microbial activity. Thus, the functional

component of the modelling framework should provide the ability to simulate oxygen transport within a given soil structure. Furthermore, the framework must contain functionality that incorporates the process of microbial respiration.

1.3 Proposed Solution

This thesis presents a modelling framework that integrates the physical and biological elements of the soil microenvironment. The requirements and design of the modelling framework are informed by a review of the relevant literature. The review, presented in Chapter 2, considers the influence of soil structure on the functional properties of a soil microenvironment. A review of work carried out in the area of structural characterisation and quantification is also presented. Further, a review of modelling techniques is presented that highlights the challenges faced in modelling complex biological systems. The detailed design specification and implementation methods used in developing the modelling framework is presented in Chapter 3. An object-oriented design strategy is introduced and it is shown that this design strategy allows the modelling framework as a whole to be systematically architectured with individual components designed, thus allowing the framework to evolve. A major consideration in the development of the modelling framework is the use of random number generators (RNG's). Random number generators are used heavily in modelling and simulation activities and can have considerable consequences on derived results. A review of the different types of RNG's commonly available is presented in Chapter 4 with details of the advantages and disadvantages of each type of RNG. A range of RNG's are tested to examine the overall randomness of the number sequences generated and a subset of these generators are also tested in a simulation environment. The results from the testing provide evidence to make an informed choice on the generator to use within the modelling framework. As this is a fundamental component of the modelling framework it is crucial that the incorporated RNG performs as expected in order to have confidence in the results obtained from any model or simulation.

A suite of metrics is introduced in Chapter 5 where each individual metric is used to quantify some element of the physical structure of soil. This suite of quantification metrics is shown to describe a range of structural properties, thus providing a practical and robust measurement tool for soil structures. The detail of each metric calculation is presented along with a discussion of how the metrics have been implemented within the modelling framework. It is also shown that these quantification metrics can be related to particular elements of soil function. Moreover, a set of CT-imaged soil structures are introduced that span a range of porosities and textures. These are used to show the validity of the suite of metrics as a tool for accurately quantifying soil structure.

Chapter 6 details a model used to simulate 3D soil structure based on training data obtained from 2D sections. The model is validated using the suite of quantification metrics to compare the simulated structures against the CT-imaged structures, from which training data was extracted. The CT and simulated structures provide the 3D physical habitat for the simulation of oxygen transport and microbial respiration in soil as presented in Chapter 7. This Chapter aims to investigate the impact of the physical structure on soil functional characteristics. The results presented also examine the relationship between structure, microbial distribution and the emergent system-scale respiration.

The concluding Chapter (8) considers the strengths and weaknesses of the work presented in the thesis, together with immediate extensions to the modelling framework. New research is outlined, within the context of the appraisal of this programme of research and with a view to future directions.

Chapter 2 – Literature Review

2.1 Soil Structure

The above-ground ecosystem depends on the ability of the soil ecosystem to provide the conditions for plant growth and animal populations. Therefore, the influence of the soil ecosystem must be considered in any management regimes of the above-ground system. Thus, an understanding of the influences of the physical structure of soil on biotic activity is sought. The physical structure of soil is comprised of two distinct phases: solid particles and void space or pores. The porous regions within the structure influence the transport of vital resources such as water, oxygen and nutrients around the microenvironment. The structure of soils, specifically the pore regions, is fundamentally determined by the chemical nature of the material such as clay particles (Marshal and Holmes, 1979). However, the biotic element also contributes to the structural properties of soil (Chenu, 1993, Feeney, et. al., 2006). The quantification of the complexity of soil structure is a significant challenge to progress in understanding the role of soil structure as a habitat for soil micro-organisms.

The investigation of structural properties of soil has, until recently, been largely based on analysis of biological thin sections of soil. These thin sections provide a 2D slice of soil that can be analysed using microscopy to identify the solid and pore regions of a sample. However, this is a destructive means of investigation, and lacks information about the 3D structure. The studies carried out by Vogel and Kretzschmar (1996) and Vogel (1997) use 2D serial thin sections and digital image processing to create a 3D representation of the soil, which then allows the characterisation the physical structure. The method used by Vogel and Kretzschmar (1996) and Vogel (1997) to capture the soil structure images result in the identification of pores greater than 30μm; pores below this resolution cannot be identified. The methodology requires image processing to be carried out to segment the digital images into a binary format (representing solid and pore states) so that quantification of the physical structure can be obtained. The methods for subsequently quantifying the structures are discussed in Section 2.2.

The recent advancement of CT technology has provided a means to capture the 3D structure of soil non-destructively and at high resolutions. The use of CT scanning technology has for the first time allowed the internal soil structure to be studied at the micro-scale and in three-dimensions. However, to date this technology is unable to detect the micro-biotic element within the soil at this scale. Perret et. al. (1999) used CT scanning and 3D reconstruction techniques to identify and quantify macro-pores in undisturbed soil cores. This study showed that this non-destructive technique of CT scanning is a useful approach to the study of the internal structure of soil samples. Perret et. al. (1999) were able to acquire an image resolution of up to 195μm, which was sufficient to allow the quantification of the macro-pore regions of the soils being investigated. The quantification techniques used by Perret et. al. (1999) are presented in Section 2.2.

Nunan et. al (2006) have shown that the use of CT technology in conjunction with image analysis techniques can result in the acquisition of high resolution structural data (c. 4.4μm). The structure of aggregates sampled from three different treatments (control, sewage sludge, and biocide treated) were investigated. There were no significant differences found in the pore geometry of the aggregates as quantified using image analysis techniques. However, at a resolution of 4.4μm a high degree of structural variation was found indicating the high spatially heterogeneity of the micro-biota habitat. Structural variation was measured using statistical analysis techniques and produced measures of spatial distribution of the pore space within the structure.

The investigation carried out by Feeney et. al. (2006) provides further evidence that CT technology can be used to study the 3D micro-organisation of the soil-root-microbe complex. The ability to obtain images at a resolution of 4.4μm allows the investigation of the structural properties of soil microenvironment at a scale relevant to microbial activity. The experiment conducted by Feeney et. al. (2006) used planted and unplanted soil cores in an attempt to regulate the bulk biotic activity within samples. The unplanted soil cores represent samples with low biotic activity as these cores contain the reconstituted arable soil that was initially

prescribed soil for all cores. The planted soil cores provide samples of relatively high and medium biotic activity levels by sampling from the rhizosphere and surrounding bulk soil respectively. All cores were incubated for 30 days, however destructive sampling took place at various points throughout the incubation period proving data on changes in root growth, soil structure genesis, water repellency and fungal biomass. Cores were also sampled at zero-day incubation to provide a base line for any changes in pore geometry to be observed. Samples from the zero-day and 30-day incubation periods were digitised using a CT scanner and image processing tools to allow the internal soil structures to be investigated in three-dimensions.

The porosity of the soil structures were calculated from the digital images and spatial correlation within the samples were analysed using semivariance analysis. The porosity and correlation lengths of the rhizosphere and bulk soil structures were shown to significantly increase over the incubation period. Further analysis carried out by Feeney et. al. (2006) through destructive sampling at points throughout the incubation period showed that large-scale stable structures were constructed within all samples, with the highest fraction of large stable aggregates occurring in the rhizosphere samples where the highest level of biotic activity was present. Water repellency and fungal biomass was shown to increase over time with significant increases shown in fungal biomass across all samples. Again the largest increase in fungal biomass was observed in the rhizosphere samples. A correlation between the fungal biomass and water repellency was also observed in all samples including the control samples, suggesting that roots do not contribute the most significant direct effect on water repellency. Instead increased water repellency was attributed to the native microbial populations. These findings provide evidence that soil ecosystem is cable of self-organising behaviour at the micro-scale with microbial communities acting on the physical structure to alter the microhabitat to more favourable conditions. This study by Feeney et. al. (2006) gives support to the theory that soil is a complex adaptive system where hierarchical organisation results in system scale emergent properties. Furthermore, it shows that the advancement in CT technology and related methodologies allow the explicit consideration of the physical structure of soil at scales relevant to microbial processes.

Historically, studies into the influence of soil structure on physical and biological processes have been conducted in a theoretical manner due to the relative difficulty of obtaining structural data in three-dimensions. These studies employ a number of different modelling techniques, such as fractals (Crawford, 1994), network models (Peat et. al, 2000), and stochastic models (Wu et. al., 2004) in which soil structure is explicitly represented in some manner and this representation is then quantified to obtain information about the structural and / or functional characteristics.

A number of studies have been conducted using network models to represent pore structure (Blunt, 2001, Vogel et. al., 2001, Johnson et. al., 2003, Matthews et. al., 2006). Network models provide a simplified representation of the pore structure using nodes as a pore volume elements and throats usually represented as cylinders to allow pore connectivity to be characterised. Models of this type can be used to predict a number of pore-level properties and are predominantly used to investigate and predict permeability and water retention characteristics of soils in three-dimensions. However, in order to construct and calibrate a pore network model, fitting parameters that relate to the particular functionality being investigated must be obtained. Obtaining such parameters can often be more straightforward than constructing the models, and because network models only produce a highly simplified representation of pore structure, they are of limited use when investigating soil structure as a habitat for micro-biota, as in this case the details of the relative distribution of pore volume and pore surface area is needed. The use of network models in predicting hydraulic characteristics of soils is discussed further in Section 2.3.

In comparison to network models, fractal based models not only model pore structure but also the solid matrix. The application of fractal geometry as a theoretical framework to represent soil structure has been employed by many researchers (Crawford, et. al., 1993b, Anderson et. al., 1998, Baveye et. al., 1998, Pachepsky et. al, 2000, Rappoldt, et. al., 2001). It has been shown that soil structures exhibit fractal-scaling properties (Dathe and Thullner, 2005). Thus fractal modelling can be used as an effective tool in soil science studies. The theory of fractal

geometry has allowed structural properties to be explicitly represented and quantified. The quantification of soil structures via various fractal measurements is discussed in Section 2.2. Furthermore, fractal geometry has provided a theoretical framework that allows relationships between physical structure and biological function to be investigated (Crawford, et. al., 1993b, Anderson, et. al, 1998).

Crawford 1994 used a fractal model to approximate soil structure and carry out an investigation into the relationship between structure and hydraulic conductivity. A random recursive lattice method based on the *n*-dimensional Cantor set (Falconer 1990) is used. The model is initiated by constructing a lattice of unit side comprised of M^d cells where d is the Euclidean dimension of the space containing the lattice; this is the primary structuring element of the lattice. The cells within the lattice are then assigned to be solid matrix by a probability p, those that are not assigned to be solid matrix are therefore assigned to be pore cells with probability $(1 - p)$. The next step is to construct a lattice of side M comprised of M^d cells each of unit side. Again, with probability p each cell is assigned to be occupied but this time the cell is occupied with a copy of the lattice constructed in the previous iteration. This process is repeated until at the nth iteration the lattice will have side M^n where up to this scale the solid matrix will share the properties of a fractal. It is important to recognize that a fractal approximation of soil structure constructed in this manner will have a lower cut-off where fractal properties are relevant, i.e. the lower cut-off is the primary structural element. The hierarchical aggregation of the primary structuring elements provides an approximation of soil structure that captures the heterogeneity of the soil matrix and tortuosity of the pore space.

Wu et. al (2004) use a Markov Chain Monte Carlo (MCMC) method to model soil structure in 2 dimensions. Because soil structure is spatially correlated it is means that the state of any structural element is conditionally dependant on the states of the structural elements in the local vicinity. This method uses a localised 6-cell neighbourhood to derive conditional probabilities from 2D biological thin sections from which to parameterise the model. With an initial estimate to begin a structure realisation, the simulated structure is built up by applying the localised neighbourhood in a raster scanning fashion and using the conditional

probabilities to derive the state of each subsequent pixel state based on the states already generated. This method is iteratively applied updating the simulated structure pixel states by successively applying the conditional probabilities until the solution converges. Wu et. al. (2004) validate this model by simulating structures with conditional probabilities derived from know 2D structures and quantifying both the simulated and real structures using four metrics: porosity, fractal dimension, variance of porosity and a semivariogram of the pore space. It is shown that the metrics calculated from the simulated structures were in agreement with the metrics calculated from the real biological thin sections.

Wu et. al. (2006) present an extension to this model that allows 3 dimensional structures to be simulated from known 2D data. The parameterisation of the model is carried out in the same manner as Wu et. al (2004) however, three perpendicular 2D slices are used in an attempt to capture the spatial correlation of the structure in all 3 dimensions. A raster scanning scheme is again used to reconstruct the 3D structure using a 3D anisotropic local neighbourhood from which new voxel states can be derived from those states already generated. Wu et. al (2006) show that this model can adequately reproduce 3D structures of relatively homogenous rock samples in terms of porosity, permeability and pore size distribution. A single simulation of a heterogeneous soil is also reconstructed by Wu et. al (2006) using this model and good results are obtained. A full description of this model is presented in Chapter 6 along with validation and results obtained from using the model to simulate a range of 3D soil structures.

2.2 Quantification of Soil Structure

To build models of how soil structure influences biological activity within the soil ecosystem the physical structure must first be quantified. Because the physical structure of soil has been shown to influence biotic activity, by quantifying the physical structure using appropriate metrics it may possible to predict the consequences for the functional characteristics of the soil. However, because the soil system is a dynamic medium, this type of quantification can only be considered as a snapshot in time of the soil microenvironment. Methods are available that describe the structural complexity of soil, mainly in terms of metrics such as porosity, connectivity and distance distributions. However, much work has also been conducted using

mathematical morphology to determine the topology of structural elements of porous media as well as fractal characterisation.

A commonly used method of quantifying soil structure is to calculate the pore size distribution (PSD). The pore size distribution can be linked to the hydraulic properties of soil and experimental and theoretical methods have been developed to measure the PSD of soils. Simms and Yanful (2004) present a discussion on the application of mercury intrusion porosimetry for measuring the pore size distribution of soils. A completely dry soil sample can be immersed in mercury, which is a non-wetting fluid, and as pressure is applied the mercury is forced through the pore networks allowing the porosimeter to measure pore sizes. However, the use of this technique to measure PSD limits the size of samples that can be analysed and the process can destroy the soil structure to a certain degree giving inaccurate results.

In contrast, Vogel (1997), Vogel (2000) and Vogel and Roth (2001) use mathematical morphology to calculate the 3D PSD of soils. They use image analysis to detect pore pathways between parallel 2D images, which results in a representation of the 3D pore structure. Vogel (1997) shows that the opening-size distribution defined in mathematical morphology can be used as a good approximation to the water retention characteristics of porous media. The opening-size distribution is calculated using the morphological operations of erosion and dilatation in conjunction with a structuring element defined over a range of radii, representing maximum size balls. The PSD calculated using the opening-size distribution is compared to the hydraulic radius calculated from water retention measurements. Vogel (1997) shows that although there are some discrepancies, this is a viable method of characterising the hydraulic properties of porous media. However, Vogel (2000) shows that network models calibrated on the moisture release characteristic can predict hydraulic properties whereas those based on the PSD cannot providing further evidence that the PSD does not fully capture the hydraulic properties of soil.

In addition, Vogel and Krestzschmar (1996), Vogel (1997) and Vogel (2001) use variants of the Euler number to characterise the topology of 3D pore networks. The Euler number is a measure of the 3-dimensional connectivity of the pore space and is based on fundamental topological properties such as number of isolated objects, number of redundant connections or loops and the number of enclosed cavities. The theoretical background relating to the Euler number and the opening-size function can be found in Serra (1982). It is shown by Vogel (2001) that the pore-size distribution and Eluer can be used as parameters to build a network model of pore structures that can sufficiently predict the hydraulic properties of the samples from which the measurements were obtained. Thus, the morphology and topology of pore networks can be considered as a governing factor in solute transport. Vogel et al (2002) used a similar image analysis technique to identify and described the pore geometry using pore-volume density, pore-surface density and the Euler number of the pore space. It is shown through the simulation of gas diffusion that these metrics can give an indication of the soil function.

In comparison Perret et al (1999) used CT scanning to obtain digital images of undisturbed soil cores. This non-destructive technique results in a complete 3D representation of the soil structure in terms of pore and solid matrix. The purpose of this study was to use the 3D images to investigate the 3D macro-pore structure of soil. The macro-pore networks are influential in transporting fluid, such as water, around the soil system. Therefore Perret et. al. (1999) developed computer algorithms to calculate both the geometry and the topology of the pore networks via a suite of metric. The metrics used by Perret et. al. (1999) were tortuosity, a dimensionless factor always greater than 1 that determines the complexity of a pore network, the 3D hydraulic radius which can be used to identify pore throats as these affect the percolation rates, the numerical density of a pore network which identifies the number of pore networks per unit volume, the coordination number which defines the number of pore throats that meet at a given point along the pore network and the genus which characterises the connectivity of a pore network. The implementation of these metric calculations requires rules defining logical relationships that act as connections between voxels to be incorporated into the algorithms. This was achieved using two logical relationship rules: nearest six neighbours

and nearest twenty-six neighbours. It was found that the nearest six neighbours criteria was computationally fast but did not identify all pore voxel connections in a single pass whereas the nearest twenty-six neighbour criteria, although not as fast did identify all clusters in a single pass. Perret et. al. (1999) showed that these metrics were useful in characterising the 3D properties of soil macro-pore networks and could be used in future studies to determine the effect of the 3D macro-pore networks geometry on solute transport.

With the ever-increasing availability of 3D soil structure data, tools that use image analysis to carry out characterisation and quantification of 3D structures have been developed. Delerue and Perrier (2002) have developed the DXSoil library specifically for use with porous media such as soil. The DXSoil library has been developed with, and integrated into the freely available OpenDX software (http://www.opendx.org); which is used for exploration and visualisation of 2D and 3D data sets. The DXSoil library consists of a number of computer programs comprised of specific algorithms that allow the extraction and analysis of the 3D pore structure of soil. However, the algorithms presented by Delerue and Perrier (2002) have been developed in a way that the geometry and topology of any 3D binary object can be identified and modelled, and in theory this should be possible at any scale. The library provides algorithms, based on a maximum ball methodology similar to that of Vogel (1997), to identify the skeleton, or medial axis, of pore networks, a local aperture map, and algorithms to identify individual sub-objects (i.e. individual pores) within the 3D pore structure. Delerue and Perrier (2002) use these algorithms as the basis of applications that calculate the pore-size distribution and retention and conductivity curves of 3D binary soil structures and show that the results obtained using the DXSoil library were comparable with those obtained experimentally.

Another software package, called 3DMA-Rock, has been developed by Lindquist et. al. (n.d) which automates the analysis of the 3D pore structure of rock. This is similar to the software developed by Delerue and Perrier (2002) in that it uses image analysis techniques to identify, construct and interpret the hydraulic properties of 3D pore structure. However in addition, the 3DMA-Rock application provides image segmentation functionality, thus allowing grey-scale

images to be analysed as well as binary data sets. This application also provides functionality that allows the construction and modification of the medial axis, or central skeleton of the pore networks. Furthermore, 3DMA-Rock can identify pore throats, pore surfaces, the complete pore-throat network and geometric characterisation in the form of various frequency distributions. The latest versions of this software have been extended to carry out analysis on the fluid distributions within 3D CT-imaged datasets. The rock samples are injected with X-ray absorbing dopants such as iodine and bromine, which increase the contrast between fluids and CT scanning of the sample can then detect where the fluids have penetrated into the pore space. This technique provides a way of obtaining penetration data of samples under a number of different fluid injections and the software is then used to examine the fluid distributions within the rock samples.

As introduced in the previous section, it has been shown that soil structures exhibit fractal-scaling properties and many researchers have used fractal models of soil structure to investigate structural and functional characteristics. Therefore, it is not surprising that a number of fractal measurements have been employed to quantify the internal heterogeneity of soil structures. The most commonly used fractal measures are the mass fractal dimension, the boundary or Minkowski dimension and the spectral or fracton dimension. The full theoretical background to these measures can be found in Mandelbrot (1983). The mass fractal dimension measures the clustering properties of a particular phase of the soil structure, i.e. either the solid matrix or the pore space can be characterised using the mass fractal dimension. The boundary of Minkowski dimension is used to characterise the interface between the solid matrix and the pore wall. The spectral or fracton dimension is usually applied to the pore space of a structure and is used as a measure of the tortuosity or complexity of the pore networks. It has been shown that a structure may exhibit fractal pore space, fractal matrix or a fractal boundary at the pore-matrix interface (Crawford et. al 1993a, Dathe and Thullner, 2005). Crawford et. al. (1993b) use fractal geometry to quantify soil structure and show that physical structure is influential in gas transport and microbial dynamics. It is found that the both the fractal dimension and spectral dimension of the pore

space are required to fully characterise the diffusive properties of the soil that influence gaseous transport through the soil system.

2.3 Impact of Structure on Function

The ability of soil to sustain microbial life is crucial to many aspects of terrestrial life on earth as the soil-microbe complex provides essential services to agriculture, waste management and the water industry as well as playing an important role in global carbon dynamics. Therefore, understanding the functional characteristics of soils is vital. The many physical, chemical and biological processes that occur within the microenvironments of soil are ultimately influenced by the geometry of the pore networks. However, the structure of both the pore space and solid matrix influences functional properties such as aeration. The process of gas diffusion, such as oxygen, into the soil matrix is influenced by the air-filled porosity. However, the complexity of the pore space, such as the distribution and connectedness of the pore networks, directly affect the availability of oxygen to the microbial populations contained within the soil matrix. The various methods of pore-space characterisation as described above can give some insight into the level of aeration of a soil microenvironment. Parts of the soil system may also provide conditions for anaerobic processes such as denitrification to occur. These types of processes are influenced by the moisture content of the soil and the distribution of distances of saturated matrix to nearest air-filled pore.

Rappoldt and Crawford (1999) investigate the anoxic volume distributions of soil using a 3D fractal model of soil structure. Volumes of a structure may become anoxic when the local use of oxygen, for example the oxygen uptake by microbes, is higher than the rate of oxygen diffusing into the volume. Rappoldt and Crawford (1999) simulate a diffusion-respiration process through a 3D random fractal lattice using a standard explicit numerical method. It was found that unconnected pores play a dominant role in the resultant distribution of anoxic volumes as when the anoxic fraction is predicted based on all pores containing oxygen the anoxic fraction can be orders of magnitude smaller than when unconnected pores are present and oxygen is absent. The anoxic fraction was also found to decrease as the characteristic

length of the diffusion process increased. In addition, Rappoldt and Crawford (2001) present a method to approximate the anoxic fraction of a fractal soil. This work shows the relation between the anoxic fraction and the distance distribution of solid voxels from connected pores when studying the local soil conditions and this approximation proves to be faster that calculating the numerical solution of a diffusion-respiration process. Both Rappoldt and Crawford (1999) and Rappoldt and Crawford (2001) highlight the important influence of the physical soil structure on gaseous transport processes.

Khalil et. al. (2004) investigate the importance of nitrification and denitrification processes as affected by the concentration of oxygen in soil aggregates. Both these processes are responsible for the production of nitrous oxide (N_2O), which is a factor in the greenhouse effect. Khalil et. al. (2004) show that denitrification increases as oxygen decreases which is expected as denitrification is mainly an anaerobic processes although some bacteria are capable of denitrification in aerobic conditions. Furthermore, nitrification rates were shown to reduced with reduced oxygen pressures again as expected due to nitrification being an aerobic process.

The contribution of soil to the global carbon budget is a current concern. As the atmospheric build-up of greenhouse gases has rapidly increased it has become apparent that sources and sinks of these gases must be accurately identified and quantified. The global carbon budget at present cannot be balanced so a "missing sink" (Tans et. al., 1990) is attributed with all the CO_2 sources and sinks uncertainties. The terrestrial carbon pool, comprised of soil and vegetation components, is the third largest global carbon pool with only the oceanic and geological pools exceeding it in size. Thus, it is important to understand the extent to which soil emissions are contributing to the atmospheric build-up of green house gases. The natural geological erosion of soil displaces soil organic carbon (SOC) from its place of formation due to wind, rain, runoff, etc. and deposits it at depressional and/or protected sites. This is a constructive process that produces fertile land that has supported civilisations for thousands of years (e.g. the valleys of the Nile). However, land use changes such as deforestation or moving from natural landscape to managed agricultural land cause accelerated soil erosion,

which degrades soil structure and diminishes soil fertility. This accelerated soil erosion changes the dynamics of the soil ecosystems and thus the emission levels will change. However, measuring emission levels from changes in land use currently lacks precision at local, regional and global scales (Lal, 2003).

2.4 Modelling and Simulation

In any discussion of quantification and prediction, a clear distinction must be made between a model and a simulation. In simple terms, a model defines the problem domain through identification of model scale, parameters, variables and mathematical construction, whereas a simulation is an individual realisation of the model with specific parameter values. Any model is at best an approximation of the real system being modelled so the modeller must compromise between perfectly simulating physical reality and producing a model that effectively approximates the physical system. There are a number of considerations that are key to producing effective models. At the core of any modelling/simulation framework is a random number generator (RNG) as this provides a means of simulating decision-making and random events. There are many types of RNG's available however it is important to note that some RNG's behave paradoxly in a non-random manner. In any modelling scenario it is imperative that the sequences of numbers generated by an RNG have proven statistical and empirical evidence that the number sequences are indeed random. A detailed discussion of the types of RNG's available and how to choose an RNG for modelling purposes is presented in Chapter 4.

Further considerations in creating effective models are the model design strategy, testing employed throughout the development of the models software and the ability to effectively communicate the purpose and results obtained from the model. The design process is a critical stage in any model development as it is at this point that the core elements of the model functionality are encapsulated. Testing of models is also essential to gain validation of, and confidence in, the model and results. The testing procedures should be rigorous and have fully documented outcomes. The validation and testing of a model ultimately results in

an iterative development of a model design and software implementation. Furthermore, communication of the model's purpose and results is fundamental, as without the ability to communicate to other researchers, industry and governmental bodies the findings from any studies using the model will be worthless.

The area of individual based modelling (IbM) can be effectively applied to the soil microenvironment as this type of modelling has been fuelled by the desire to understand the natural complexity found in many biological and non-biological systems and importantly how the large-scale system complexity emerges from variability and adaptability of individual component parts. An IbM is designed from the bottom up, where properties of, and interactions between, the components at the bottom level of the system are encapsulated into individual functional object types with model input parameters defined at this level. This design methodology allows for many component types to exist within a model and importantly, allows many components of the same type to exist and with different property values thus introducing variability in individual behaviour. From this, system-scale properties emerge as a consequence of the interactions between the component parts. This design strategy can be applied to the study of complex adaptive systems as discussed Chapter 1 (Crawford, et. al., 2005, Levin, 1998).

The individual based approach to modelling provides a common language by which a model can be developed and communicated, in that it provides a way of thinking and talking about a model that is relevant both to the real world and to the software design. An object-oriented design strategy is commonly used in the design and implementation of an IbM as object-oriented design and programming approach naturally encapsulates the state, behaviour and interactions of individual model components. The software implementation of a model is fundamentally important, especially when implementing an IbM as the software is the model. In the words of Grimm and Railsback (2005), an "IbM's software must provide a laboratory for experimenting on the model". Therefore, any models software must be tested and validated against known theory and experimental data to provide evidence that the model implementation, i.e. the computer code, is correct.

Kreft et. al (1998) present an object-oriented simulator, BacSim, that is used to model bacterial colony growth using an individual based methodology to simulate diffusion, reaction and growth processes. The fundamental question being studied in this work is whether it is possible to create a macroscopic world from data on microscopic entities? The model presented by Kreft et. al (1998) is an extension of Gecko, a simulator that has been sued to study ecosystem dynamics (Schimtz and Booth, 1997). Gecko itself is built using the Swarm toolkit which has been developed by Santa Fe Institute, New Mexico, USA (Minar et. al., 1996) with the aim of providing a software platform, comprised of a number software libraries, that are applicable across a variety of fields where complex adaptive systems are of interest.

The BacSim simulator models each bacterial cell as an individual object within the computer simulation, where each cell has the same potential state and behaviours. The bacterial characteristics (i.e. behaviours) that BacSim models are substrate uptake, metabolism, maintenance, death and cell division, these processes are carried out using parameter values and algorithms for individual cells that are found in relevant literature. Each cell occupies a single lattice element on a 2D spatially discrete grid and substrate diffusion is also simulated within this grid. Each spatial lattice element may contain substrate and this substrate is available to any bacterial cells that also occupy this lattice element, thus allowing each bacterial cell to evolve within the limits of the available resource. The evolution of the system is carried out over discrete time-steps where each bacterial cell repeats their functionality once during each time-step. Kreft et. al (1998) validate each specific element of the model for both functional and numerical correctness against known theoretical and experimentally obtained results. It is shown that both the physical and biological element of the system are integrated into the model. Kreft et. al. (1998) show that the individual based approach to modelling complex adaptive systems, such as bacterial colonies, is viable. This approach allows knowledge of the functionality of individual system components to be encapsulated and used to simulate the behaviour of the system as a whole.

Ginovart et. al. (2002) use a similar methodology and also present an individual-based model, INDISIM, that simulates the growth of bacterial colonies. Again, this model is discrete in both space and time and is capable of simulating the global behaviour of bacterial colonies from a few input parameters that are defined at the level of the individual bacterium. INDSIM controls a group of bacterial cells and the state of each bacterium along with the behaviour and interactions, such as the movement, nutrient uptake, metabolism and reproduction of each bacterial cell in relation to its spatial location and resource availability. The model is used to simulate specific experiments that were also carried out under laboratory conditions and the simulated results show qualitative comparison with the experimental results. In addition, Ginovart et. al. (2005) extend the model and present INDISIM-SOM which allows the study of microbial activity in soil in terms of carbon and nitrogen mineralization and the nitrification process. This work is a natural extension to Ginovart et. al. (2002) and shows that the individual based approach to modelling is versatile and adaptive. The simulators presented by Ginovart et. al. (2002) and Ginovart et. al (2005) incorporate both biotic and abiotic elements and therefore enables the model to account for spatial heterogeneity of bacteria cells and resources. The INDISIM-SOM simulator models the dynamics and evolution of two different types of microbial cells: decomposer micro-organisms and nitrifier bacteria in relation to resource availability, in that it takes into account the role of carbon and nitrogen in the life cycle of these microbial cells. These type of individual based studies require experimental data to be available for calibration, validation and parameterisation. Ginovart et. al (2005) used experimental data obtained from the incubation of two Catalonian soils for calibration of the model. However, the model does not explicitly account for the spatial heterogeneity of the physical structure of soil.

2.5 Summary

The review of the relevant literature presented in this Chapter has highlighted the challenges faced in the investigation of the soil-microbe complex. The literature shows the importance of incorporating both the biotic and abiotic elements of the soil microenvironment into a model as the spatial heterogeneity of both these elements affects the global functional properties. All

the literature reviewed above demonstrates the significance of the physical structure of soil in regulating the transport of resources and microbial activity within the soil system.

Due to the progress in CT technology, it is now possible to obtained three-dimensional structural data at the micro-scale (Perret et. al. 1999, Nunan et. al, 2006). Therefore, the challenge is to relate the 3D structural characteristics to the functional processes that occur at this scale in soil. Feeney et. al. (2006) showed experimentally that soil-root-microbe system is adaptive and that the micro-structure of soil is affected by resident microbial activity. However, no explicit link between structural properties and microbial activity was made. Models of soil structure have also been useful in gaining insight and understanding into the role of structure in the regulation of soils functional characteristics. Network models have been used to simulate the hydraulic properties, such as water flow and retention by represent the connectivity of the pore structure or soil in terms of pore volumes and connecting throats (Blunt, 2001, Vogel et. al., 2001, Johnson et. al., 2003, Matthews et. al., 2006). Fractal models have proved to be a useful 3D approximation of soil structure that also allows functional properties to be investigated (Crawford, et. al, 1993b, Anderson et. al, 1998).

However, to explicitly relate structure to function, the structural characteristics must be quantified in a manner that facilitates this relationship. A number of software packages have been developed to allow the quantification of soil structure such as DXSoil (Delerue and Perrier, 2002) and 3DMA-Rock (Lindquist et. al. (n.d)) but these toolkits are mainly concerned with hydraulic properties of soil. To date, no such toolkit exists that facilitates the quantification of structure in terms of aeration properties even with the availability of methods and metrics to do this (Perret, 1999, Rappoldt and Crawford, 1999, Vogel, 2002,)

Due to the complex nature of the soil microenvironment as detailed in the above review of the literature, the individual based approach to modelling this system is natural design methodology. The individual based modelling approach has been shown to be effective in integrating abiotic and biotic elements of a system (Kreft et. al, 1998, Ginovart et. al, 2002) and controlling the system-scale processes. Therefore, the individual based methodology will

contribute largely to any further investigation into complex adaptive system such as the soil-microbe microenvironment.

Chapter 3 – The Modelling Framework: Design and Implementation

3.1 Introduction

Chapter 2 highlighted the importance of software in modelling and especially in individual based models, where the software must act as a virtual laboratory for experimenting with the model. This chapter introduces an object-oriented strategy that is used to design and implement the modelling framework that facilitates the investigation of the soil microenvironment. The object-oriented design strategy employed provides a way of thinking about the problem domain that reflects both the real world and the implementation of that world in the software-programming environment. The strategy used to develop this modelling framework is one where the whole system is systematically architectured and then the specific components are designed using an object-oriented approach. This means that as the system evolves the integration of new features is made easier than if using more traditional approaches such as procedural programming, which was traditional in the 1980/90's. The implementation of computer-based models using a procedural programming approach simply consists of coding algorithms. However, this type of modelling strategy is difficult to maintain in terms of extending functionality and integrating new components as the model develops. The procedural approach is also limited in terms of its expressive power for encapsulating state and behaviour of particular model components. As the modelling of the soil-microbe system is essentially about modelling the behaviour and changes in state then the object-oriented paradigm is more appropriate.

This Chapter describes the design and implementation of a software architecture developed to provide a complete, reusable and extendable modelling framework. The object-oriented (OO) methodology used in the design and implementation of the modelling framework is discussed in Section 3.2. Using the OO strategy the model design is built into the implementation, where class libraries encapsulate specific model components including both the state and behaviour of these components. The design specification detailing the core

requirements of the modelling framework can be found in Section 3.3 and the system architecture is detailed in Section 3.4. It is shown that the OO approach to modelling, especially biological modelling, allows the components of the system to be described at an individual level but leads to prediction of the behaviour of the system as a whole. Thus, a modelling framework is presented that allows any emergent properties or self-organisation within the system to unfold during any simulation realisation.

A further aspect of any modelling framework is how to parameterise the model domain. This can often be ad-hoc and hence can be difficult to identify, maintain, verify and document. However, this project has incorporated a novel system called Plang, which provides a method to formally define, maintain and document model parameters. The key features of Plang are introduced in Section 3.5 with an overview of how this will be used for experimentation with the model framework. The testing and implementation methods used are presented in Section 3.6 and Section 3.7 provides a discussion on this object-oriented design methodology for use in biological modelling.

3.2 Design Methodology

An object-oriented (OO) strategy is used to design and implement the modelling framework. The framework can be thought of as a general design for implementing any number of models based around the soil-microbe system. A framework usually consists of a number of code libraries that describe a collection of reusable components that encapsulate the state, behaviour and interactions of individual entities of the model domain. At the highest level are the executable applications that provide an interface to the end user as a means of accessing the underlying functionality. The user interface level of an application may access the functionality provided by any number of library files. A library encapsulates a set of specific model components that provide complete functionality for a particular part of the model framework.

The components contained within a library are called classes and the set of classes make up a reusable and extendable architecture. Each class describes the states and behaviours that an object of that class will exhibit and in that sense a class defines a model component type. An object is considered to be an instance of a particular class / type and these exist as individual entities within a model realisation. An object can be related to agents discussed in the literature regarding the modelling of complex systems (De Wolf, et. al., 2004; Kreft, et. al., 1998). The state of an object is defined in terms of a set of attributes or properties, that is a set of data values. The behaviour of an object is defined in terms of a set of methods that provide that behaviour. The terms state and attributes/properties, and behaviour and methods are used synonymously throughout the remainder of this Chapter. It must be recognised that attributes and methods for a particular type of object are defined at the class level. Thus, any component type within the model domain is defined only once within the framework, but the many individual entities of this type are created as instances (objects) of this type.

The OO approach to modelling complex systems uses the concept of abstraction and inheritance to describe the level of detail of each component of the model domain. Identifying objects as belonging to the same class reflects the level of abstraction, or generality within the system, thus objects can be classified at different levels of detail. Therefore, if the objects are classified at different levels of detail, the classes can be put into a classification hierarchy where the most abstract class is at the top, and is termed the superclass and the more detailed classes below are each a subclass. Furthermore, the inheritance within a class hierarchy means that all subclasses inherit, or contain all the characteristics of the class above. In addition, the subclass can also define additional state and behaviour characteristics specific to itself.

The OO methodology provides a formal yet flexible platform on which to build a software architecture to facilitate the modelling of the soil-microbe complex. The development of our modelling framework using this methodology means that framework can evolve and expand, as more functionality is required. The initial step in the design process is to identify the libraries and classes required, with each class encapsulating all the functionality required by a

single model entity. The inspiration for class designs comes from the problem domain being modelled, in this case the soil-microbe system. Thus, the appropriate elements of this complex system must be identified and defined. Therefore, the OO strategy allows both the design and implementation of the framework to be based on real physical entities such as soil structure and soil microbes but also allows for more abstract entities required within the framework to be modelled such as random number generators or the process of oxygen diffusion.

3.3 Design Specification

This Section sets out the core requirements for the modelling framework, these has been informed from a review of the relevant literature as presented in Chapter 2. A key objective of this research programme is to investigate the impact that the physical structure of soil has on the functional characteristics, specifically microbial activity. Therefore, the modelling framework must provide a virtual laboratory that facilitates investigation, experimentation and analysis of the soil microenvironment. To this end, the core requirements for the modelling framework emerge and are presented in Table 3.1. The core requirements of the modelling framework are specific in terms of the required overall functionality and the architecture presented in Section 3.4 is the general model design used to implement these requirements. The subsequent Chapters of this thesis provide explicit detail relating to each of the requirements stated in Table 3.1.

1	The 3D structural heterogeneity of soil must be explicitly represented
2	Tools must be provided that facilitate the quantification of 3D soil structure
3	Facilities that allow the 3D spatial distribution of microbes to be defined
4	The ability to simulate oxygen diffusion within the 3D structure
5	The ability to simulate microbial respiration
6	Functional characteristics / traits of all model components must be defined at the individual level
7	Must allow the ability of system-scale properties to emerge as a consequence of

	individual trait values
8	The framework must provide mechanisms to report on quantification / simulation activities
9	The framework must be accessible to users with no programming knowledge
10	The framework must be reusable, extendable and easily deployed

Table 3. 1 - Core requirements for the modelling framework

3.4 System Architecture

The design methodology described above has been employed in the design and implementation of a modelling framework that meets the core requirements detailed in Section 3.2. The units of deployment that make up the complete modelling framework are shown in Figure 3.1. The two executable applications and a five class libraries have been developed so that each provides specific, complete, reusable and extendable functionality. It can be seen that the lowest level libraries provide functionality to both the application layer and the middle layer libraries; these low level libraries provide core functionality to the modelling framework. Nevertheless, the nature of research is such that the exact detail of the interactions and processes encapsulated within these class libraries is not fully understood at the start of the framework design and so the development of each class library is an iterative process. As understanding of the problem domain increases the classes within each library can be amended and/or extended to reflect the required functionality. Thus, the architecture described in this section provides the current design.

3.4.1 Soil Architecture Model (SAM)

The Soil Architecture Model (SAM) is an executable application and provides one of the modelling framework user interfaces. A key consideration in the modelling framework design was the ability for the underlying functionality of the framework to be accessible to other researchers, either with or without the technical knowledge of the model or programming

environment. Therefore, the SAM application consists of Windows based graphical user interfaces (GUI's) that provide the end-user with access to the functionality contained in the lower layers of the framework. GUI's are a familiar and intuitive way for all users to interact with the system, including those with only limited knowledge of the modelling framework.

This application provides an interface to allow the user quantify 3D soil structures via a number of metrics (Chapter 5) and simulate oxygen transport and microbial respiration within a given structure (Chapter 7). However, due to the object-oriented design and implementation methods used in this project, the executable applications are responsible only for facilitating user input. The code for executing the core functionality of the SAM application is encapsulated in the Structure Analysis library (3.4.3), thus decoupling the input and processing aspects of the software.

3.4.2 Pore Architecture Model (PAM)

The Pore Architecture Model (PAM) is the second executable application developed as part of the modelling framework. This application is the implementation of the model presented in Chapter 6, which uses 2D data to simulate the 3D micro-structure of soil. Again, this executable application provides the user with GUI's to lead the user through the modelling process and the core functionality of the model is separately encapsulated in the Soil Builder library.

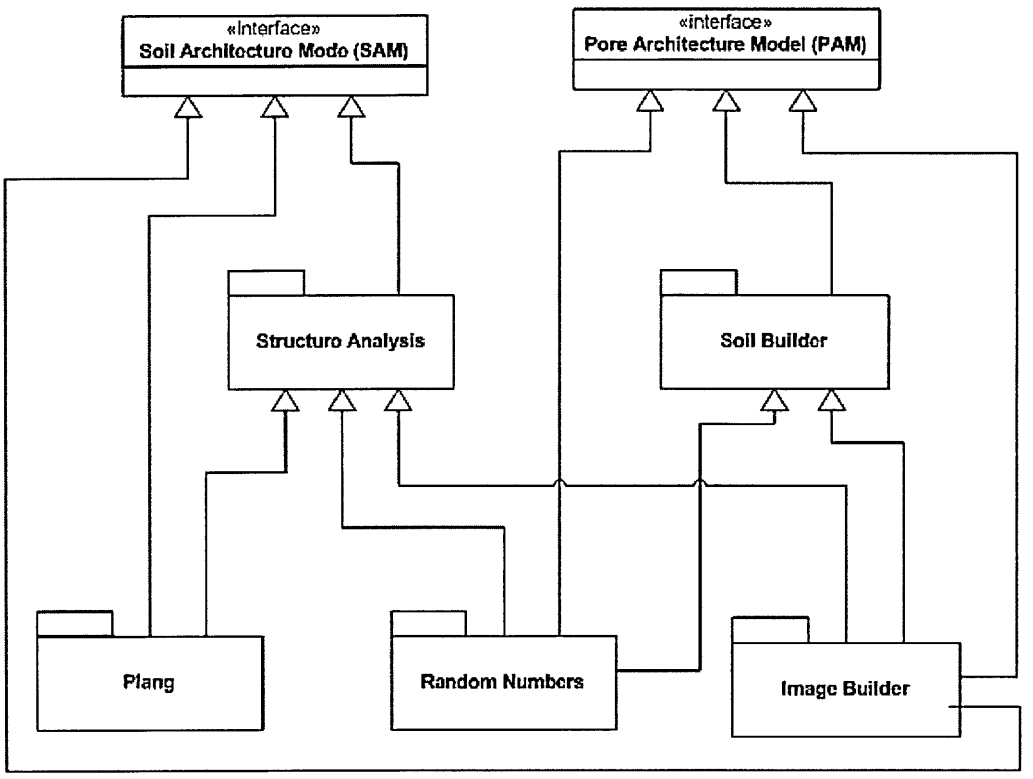


Figure 3. 1 - Units of deployment making up the complete modelling framework

3.4.3 Structure Analysis Library

This library models the soil-microbe complex explicitly in terms of entities and processes found in nature and provides the core functionality to the SAM application (3.4.1). The classes encapsulated in the Structure Analysis library are shown in Figure 3.2 and these classes provide a design and implementation necessary to meet requirements 1-8 as detailed in Table 3.1. The individual class designs shown in Figure 3.2 show the attributes and methods that are publicly exposed, i.e. accessible to an executable application or other library file. These classes require many more properties and methods to execute the necessary calculations. However, these remain private to each class and can only be accessed through the embedded functionality within the public attributes / methods shown in Fig. 3.2. The class designs embedded within the implementation include the data types of the attributes for each class. These data types are identified in Figure 3.2. The data types are an essential part of the design as these provide the means of encapsulating the state of the system. For example

a Structure object has a porosity that is represented by a double precision number indicating the percentage of volume that is pore.

The initial inspiration for the design of the Structure and Voxel classes came from the area of Computed Tomography (CT) as some of the soil structures available for study were captured in this manner (see Chapter 5). The Structure class is the main object type within the Soil Analysis library and only one object of this type is created during a particular SAM simulation. The state of a Structure object is determined by the states of the Voxel objects contained within the 3DVoxelArray attribute shown in Figure 3.2. This three-dimensional array of Voxel objects maintains the state of a structure as a whole but at an individual level, in that each Voxel object maintains its own state as defined by the attributes shown in Fig. 3.2. It can be seen that each Voxel object maintains its own location within the 3D grid via the Row, Col and Layer attributes. Making Voxel objects self-referential is advantageous for the implementation of the algorithms used in the quantification of the structures (see Section 5.4.2). The Structure class provides a number of methods that contain the algorithms to carry out quantification and simulation functionality, this includes setting object attribute values on completion of method. Each method also produces a text file with results of the quantification or simulation activity.

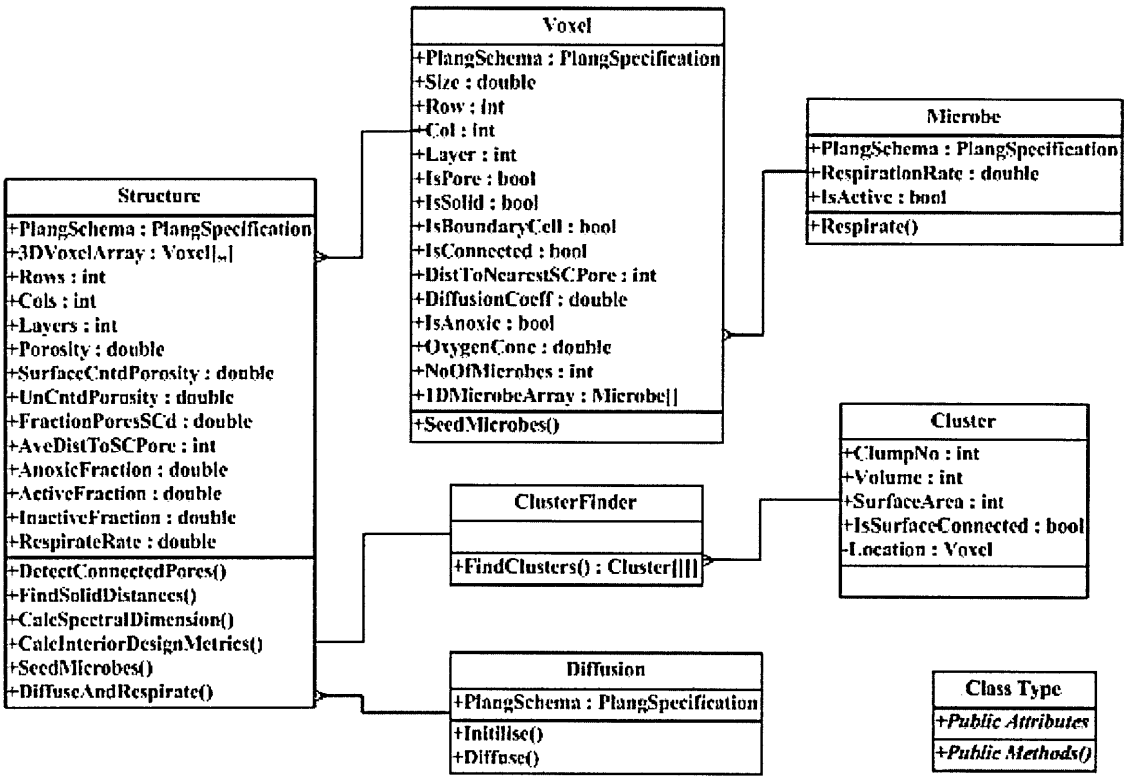


Figure 3. 2 - Class designs used in implementing the Structure Analysis class library

The relationships between the Structure Analysis classes are also shown in Figure 3.2 and these are another key element of the design and implementation. As discussed above a Structure object is ultimately described by the 3D array of Voxel objects, which are self-referential. It can be seen from Fig. 3.2 that each Voxel object contains a collection of Microbe objects (1DMicrobeArray attribute) and because each Microbe object belongs to a specific Voxel object, the behaviours the microbes can exhibit are controlled by the state of the Voxel in which it resides. The core functionality provided by the Microbe class is the process of respiration that is carried out by an individual microbe and this process is encapsulated within the Respirate method, the exact specification of this process is detailed in Chapter 7.

The Diffusion class is slightly different in that it encapsulates the numerical method by which to simulate gas diffusion within a 3D grid. An instance of this class is instantiated within the DiffuseAndRespirate method of the Structure class, thus the Structure object has complete control over the process. However, with the functionality of a diffusion process encapsulated within its own class, it is possible to instantiate more that one object of this type, thus

providing the ability to concurrently run any number of diffusion processes within a single simulation.

A further aspect of the design of this class library is way in which parameters are defined. The parameters and associated values required by each class are defined outside the modelling framework using a Plang specification file (see Section 3.5) and it is only at run-time that the parameters are loaded and assigned to the particular object through the PlangSchema attribute. This is another mechanism by which input is decoupled from process within the modelling framework. This is particularly useful in a modelling scenario where parameters values are likely to be altered on a regular basis. Plang facilitates this with ease, allowing parameter values to be changed outside the modelling framework thus avoiding the need to amend and recompile source code. The Plang system provides further stability and flexibility in both the design and implementation of the modelling framework.

3.4.4 Soil Builder Library

The Soil Builder library encapsulates the specific model components that are the design and implementation of the Pore Architecture Model presented in Chapter 6. The classes making up this library can be found in Appendix I. This library is different from the others in terms of the type of classes designed as each class encapsulates a more abstract mathematical entity.

3.4.5 Image Builder Library

The Image Builder library has been developed to allow the modelling framework to process, manipulate and store 2- and 3D structural data sets. A specific file format (.str) has been used to store structures within this modelling framework and this library provides functionality to convert data sets obtained from CT-imaging into this format. Image Builder also provides functionality to extract volumes or slices from 3D structure files. The classes encapsulated by the Image Builder library can be found in Appendix 1

3.4.6 Random Numbers Library

The Random Numbers library again provides core functionality to the modelling framework. The Random Numbers library has been developed to provide the modelling framework with a proven random number generator (RNG), as this is core to many parts of the modelling framework. The choice of RNG to use in model simulations is vitally important as bad RNG's can result in wrong results. A review of different types of RNG's and considerations for their use for simulation purposes are presented in Chapter 4.

3.4.7 Plang Library

The Plang (Parameter Language) system, developed by Milne (2006), provides a formal way of identifying and maintaining model parameters within this modelling framework. Plang was initially developed with individual-based biological systems in mind thus the system enhances the modelling framework developed here, further detail can be found in Section 3.5. The Plang system enables the definition of model parameter specifications, their verification and provides an application programming interface (API) that allows the user application to access the Plang specifications programmatically. A full system overview and examples of Plang specifications are presented by Milne (2006) at http://a510690.studentweb.abertay.ac.uk/Plang/index.html.

3.5 Parameterisation of the Model Domain Using Plang

The parameterisation of a model domain can often be ad-hoc and hence parameters can be difficult to identify, maintain, verify and document. This section will show how Plang aids and enhances the parameterisation of the soil-microbe system within the modelling framework. Due to the investigative nature of this project, the parameters that drive the processing of the simulations may be altered on a regular basis. This allows any emergent behaviour that may arise within the system to be identified. The Plang system provides a means of defining, altering and documenting parameters and specific parameter values which is vital to the modelling framework. The novel aspect of Plang is that it allows all parameters to be defined

outside the main application. This means that a program does not have to be changed and recompiled each time a parameter value is changed. The Plang system requires model parameters to be defined using a Plang specification. This is a text file written in the Plang specification language and it is only this file that needs to be altered when a parameters' value is changed. The Plang specification files are also a useful tool in documenting and auditing simulation and testing activities.

Each Plang specification packages parameters into Schemas that represent categories or types within the model domain. Each Schema contains a collection of parameter specifications. An example of the Plang specification required as input to the SAM application is given in Figure 3.3. It can be seen that there is a Schema defined for each of the class types defined in the object model presented in Figure 3.2. The definition of parameters by object type using Plang further strengthens the model design as the attributes of any class are defined by the parameters encapsulated by a single schema. Each individual parameter specification within a Schema consists of a name, type and a default value. These parameters define size and scale of model components such as voxels, and can also provide specific trait values, such as the respiration rate of microbes. However, Plang also provides functionality to define a parameter by a probability distribution, a range or closed set distribution. The default value for a parameter can either be a specific value or a sample from the defined distribution. This way of defining parameters provide the flexibility to incorporate data obtained from experimentalists. The Plang system also provides an application programming interface (API), which enables the implementation of the parameters specification in to the model framework. It can be seen in Figure 3.2 that each class has an attribute of type PlangSpecification and it is through this attribute that an object can access the specific parameter values.

```
Author J. M. Blair
Date 25/9/06
Defines the default parameters for the soil-microbe model
-----------------------------------------------------

Structure Parameter Schema

        DampingLayers is integer default = 1
          defines the number f damping layers surrounding the structure ---

        SpecDim_Walks is integer default = 1000
          defines the number of walks taken for calulating the spectral dimension ---

        SpecDim_Start is integer default = 1
          defines the walk number to start at ---

        SpecDim_Steps is integer default = 60000
          defines the maximum steps taken in any random walk when calc'ing the spectral dimension ---

        FixedBiomass is integer default = 100000
          defines the number of microbes to seed in a structure ---


        MicrobeDistanceProbabilities() is real from 0 to - default = (0, 1.0, 0.5, 0.1, 0)
          defines the number of microbes in each voxel distance class see below ---


        WaterFilledPoreSpace is real from 0 to 100 default = 0
          defines the percentage of pore space that is water filled ---


End Schema


Voxel Parameter Schema

        VoxelSize is real default = 4.5
          defines dimension of a voxel, voxel is of volume VoxelSize^3 (m)---

        EdgeProbability is real default = 1.0
          defines the probability of an potentail active voxel having a bug  (i.e. being adjacent to a pore voxel, or a solid voxel)---


        NumberOfBugs is integer from 0 to 3 default = 1
          defines the number of bugs a voxel contains ---


End Schema


Microbe Parameter Schema

        RespirationRate is real from (0.1, 0.25, 0.5, 0.75, 1.0, 10.0) default = 10.0
          defines respiration rate of a bug (mols/voxel/s) ---

        Mobile is boolean default = false
          defines whether a bug is mobile ---

End Schema


Diffusion Parameter Schema

        OxygenSupply is real default = 1.0
          defines the concentration of oxygen in the supply layer (mols/voxel) ---

        Dpore is real default = 1.0
          defines diffusion coefficient in pore space (m^2/s) ---

        Dsolid is real default = 1.0
          defines diffusion coefficient in solid space (m^2/s) ---

        h is real default = 0.01
          defines  change in x, y, z, in diffusion eqn (m) ---

        k is real default = 0.00001
          defines the timestep size (s) ---

        rjac is real from 1 to 2 default = 1.0
          defines the rjac values for SOR algorithm ---

        SteadyState is boolean default = false
          defines whether to diffuse to steady state if true, otherwise use TimeSteps below to indicate legth of diffusion ---


        TimeSteps is integer default = 10000
          defines the number of timesteps diffusion and respriation processes carry out ---


End Schema
```

Figure 3. 3 – Example of Plang Specification file

3.6 Implementation and Testing Strategy

The implementation of the object-oriented design described above has been carried out using the C# language within the Microsoft .Net framework. This platform provides an Interactive Development Environment (IDE) on which to build the executable applications and library files detailed in Section 3.4 and Appendix I. The Microsoft .Net framework is a fully object-oriented platform that is naturally suited to developing individual based models like that presented in the previous Sections. The implementation of the system architecture into computer code and the subsequent validation of the code is a crucial phase in the development of the modelling framework. Without suitable testing at all stages of development, and documenting of the testing, there cannot be sufficient confidence in any results obtained. Therefore, it is imperative that a methodical, systematic and detailed testing strategy is employed throughout the development, and maintenance, of the modelling framework.

The testing and validation carried out during the software development must establish the quality, performance and reliability of the modelling framework at all levels of the architecture. To verify the modelling framework as a whole, each class must be tested independently to verify the correctness of both the class design and numerical calculations. Furthermore, a number of different types of testing must be carried out during software development to ensure accuracy of both the overall design implementation and results generated. In order to fully document all aspects of testing, the computer code must be completely commented, both in terms of how, computationally, an operation is performed but also why it is carried in the particular manner, both in terms of the code but also the underlying biological, chemical or physical reason for carrying out such an operation. This is important for testing, maintenance and collaboration purposes and results in the model design being completely embedded within the implementation. With code fully documented, code reviews should be undertaken by someone who is familiar with the model design but is not involved in the actual coding process to detect any major programming errors or code design issues. Throughout the course of this research programme reviews were performed by a number of researchers on

small pieces of code that carry out a specific function. Thus the reviews were completed with relative speed and proved to invaluable in detecting design and coding errors at an early stage in the development process.

In addition, it is essential to compile test input data that has know results and therefore can be used throughout the development cycle to test code changes made at any level within the framework. Test input data should force a wide range of conditions to be tested within the model, including extreme inputs as this provides justification and evidence that the model implementation is correct. The incorporation of the Plang system into the modelling framework has proved a valuable tool in documenting test input data as well as driving the model. In addition, the use of the Plang system in this model has formalised the parameterisation process, thus providing a reliable and intuitive means of testing and documenting parameters, both for testing purposes and in general use of the modelling framework.

Visual testing of results is also a quick and intuitive way of inspecting the models behaviour and is particularly useful in detecting any errors in expected overall trends of simulations. The way in which testing is applied during software development is dependant on the operations and functionality being tested. Therefore, each subsequent Chapter contains a section detailing the context specific testing carried out on random number generators (4.3), the quantification metrics (5.4.3), the pore architecture model (6.2) and the diffusion process implementation (7.2.3). The SAM and PAM executable applications are responsible for facilitating user input. Therefore the GUI's were systematically tested to ensure that the user is only presented with options that are logical and timely. The GUI's are updated to allow users access to a new piece of functionality elsewhere in the framework. Each time an update is made all form opening and closing operations are checked together with form option availability. These applications are not responsible for checking the values of any parameters entered; this is carried out within the class constructor method to which the parameter is passed and will therefore be tested during class implementation.

3.7 Discussion

The software architecture presented in this Chapter has been designed and implemented using an object-orient strategy. The use of the OO approach has resulted in a reusable, flexible and extendable modelling framework. The class architectures provide the full and current description of the model design and implementation that meets all of the core requirements detailed in Table 3.1. It is shown that the modelling framework developed and presented in this Chapter is a 3D spatially explicit individual based model of micro-organisms in soil.

The OO approach provides a number of advantages in the development of the modelling framework over traditional programming techniques. Firstly, the design and implementation resembles the system being modelled and this aids the communication of the model design and results. In addition, this approach provides greater protection of data as all data are encapsulated within specific individual objects, thus each object controls the manipulation of its own data and hence state. Moreover, the encapsulation of all code within objects avoids the risk of any code changes affecting unrelated data. The object-oriented design strategy also provides flexibility in process control, as objects provide a natural means of passing execution control between model components.

Due to the hierarchical organisation of the model components, evolution and maintenance of the modelling framework is straightforward. This OO approach facilitates revision, extension and adaptation of the embedded functionality with relative ease, as any changes, in theory, should be made within a single class or class library. A change to the required functionality of a model component requires the code in only one class to be amended. Furthermore, the extension of the framework to include more model components simply requires a new class to be added to the relevant class library.

The OO design and implementation of the modelling framework coupled with the Plang system permits documentation and reproduction of simulation activities which is essential not only for testing purpose but also for communication of model results. The use of Plang within

the modelling framework formalises the parameterisation of the problem domain and decouples parameters from process. Moreover, Plang facilitates communication between modellers, theoreticians and experimentalist.

The modelling framework as presented in this Chapter shows the current architecture of the system. However it is envisaged that this software architecture will be continually reused, amended and extended to suit the needs of any future investigations into the soil-microbe complex.

Chapter 4 – Random Number Generators in Modelling

4.1 Introduction

Random number generators (RNG's) are used in a host of applications' but what makes a random number generator 'random'? This is a rather fundamental question, as no random number generator is genuinely random. L'Ecuyer (1998) stated, "What we call a random number generator is actually a program that produces, once its initial state is chosen, a deterministic and periodic sequence of numbers." From this definition any random number generator is in reality a pseudo-random number generator. However, these will be referred to as random number generators throughout the thesis.

This Chapter presents an overview of random number generators and their use in modelling and simulation. The production of good quality random numbers is a core element of the modelling framework presented in the previous Chapter with random numbers also playing a fundamental role in the Pore Architecture Model discussed in Chapter 6. Therefore, it is crucial that the RNG implemented within the modelling framework is producing good quality random number sequences. Otherwise we cannot have confidence in the results of any simulations.

Random numbers are used for many purposes in computing, from serving as data to test code through to decision making in computer games. The area that is of prime interest here is that of simulating real-world events that involve a probability, otherwise known as Monte Carlo simulations. This type of simulation is used both in the calculation of some of the metrics introduced in Chapter 4 and is core to the Pore Architecture Model. The Monte Carlo method of simulation allows many problems to be investigated and solved that otherwise would have no possibility of solution. The quality of random numbers used in these types of stochastic simulations can have significant effects on the results. Thus, the choice of random number generator is important and an understanding of the properties of different generators should be considered before implementation of an algorithm.

The following section provides an explanation of the different mathematical techniques for generating random numbers with some advantages and disadvantages of the particular generators. Section 4.3 details the testing employed to produce statistical support for a generator and Section 4.4 discusses some of the more subjective aspects that should be considered when choosing a random number generator. Section 4.5 introduces the types of random number generators investigated in this project and Section 4.6 presents results obtained from the testing of these generators using the Diehard suite of statistical tests.

4.2 Types of Random Number Generators

There are a number of different types of random number generators available for use and the following provides an outline of the most commonly used across a wide spectrum of applications including computer simulation, numerical analysis, random search optimization and computer games. The random number generators discussed below all have their own advantages and disadvantages, but in general all random number generators should possess the properties of long sequences of un-repeating random numbers, good structure, efficient computation algorithm and easy implementation along with some proven test results both theoretical and empirical (Hellekalek 1998, Knuth 1981, L'Ecuyer 1998)

4.2.1 Linear Congruential Generators

One of the most commonly used random number generators is the linear congruential generator (LCG). The algorithm below is used to generate a linear congruential sequence:

$$x_n = (ax_{n-1} + c) \bmod m \qquad (4.1)$$

where the coefficient *a,* additive term *c,* modulus *m* and the seed x_0 are parameters to be chosen. The mod *m* part of the recurrence yields the remainder of $(ax_{n-1} + c)/m$.

An LCG of this form is denoted by LCG(m, a, c, x_0). The random number on the unit interval is obtained by:

$$u_n = \frac{x_n}{m} \qquad (4.2)$$

Knuth (1981) shows that all sequences having the form $x_{n+1} = f(x_n)$, like that of the LCG, "get into a loop". This endlessly repeated sequence is a common property of LCGs, and in-fact all the generators that follow, and is called the *period*. The most useful sequences will be those with the longest period (Knuth 1981, L'Ecuyer 1998).

Some well know LCGs are the Ansi-C system generator LCG(2^{31}, 1103515245, 12345, 12345), the "Minimal Standard" generator LCG(2^{31}-1, 16807, 0), the infamous RANDU which Knuth calls "really horrible" is LCG(2^{31}, 65539, 0) and NAG's LCG(2^{59}, 13^{13}, 0, 123456789(2^{32}+1)) (Hellekalek 1998, Janke 2002)

These basic linear congruential generators are the most commonly used partly because the theory behind them is well known, but also because they are relatively easy to implement in languages such as C and Java. The LCG family have practical advantages such as speed, ease of implementation, portable code with parameters and test results that are widely available.

One disadvantage of the LCG's is that they tend to have a relatively short period, for example the "Minimal Standard" generator of Park and Miller (1988) has a period of 2^{31} and this can be exhausted quickly when carrying out Monte Carlo simulations (Janke 2002). A further drawback of this type of generator is that of the lattice-structures that are formed from D-dimensional vectors (x_1, x_2, ..., x_D) of consecutive random numbers which all lie on a small number of parallel planes, thus resulting in sequences of numbers that do not cover the entire state space available (see Hellekalek 1998, Figure 2). The spectral test is used to calculate "figures of merit" for each RNG, which is the size of empty space between each slice of the

lattice (Hellekalek 1998, Knuth 1981, L'Ecuyer 1998, L'Ecuyer and Andres 1997, Sobol and Levitan 1999). A number of tests, including the spectral test will be discussed further in Section 4.3.

4.2.2 Multiple Recursive Generators

The *multiple recursive generator* (MRG) is defined by the recurrence:

$$x_n = (a_1 x_{n-1} + ... + a_k x_{n-k}) \bmod m \qquad (4.3)$$

where then *modulus m* and *order k* are positive integers, the *coefficients* a_i belong to Z_m = {0,1, ..., m-1}. A seed value x_0 must also be selected to initiate the random number sequence. Again, to obtain a random number on the unit interval the normalisation of 4.2 is used. With only two non-zero a_i's and a prime *m* the sequence has maximal period $\rho = m^k - 1$ (Knuth 1981, L'Ecuyer 1998) giving:

$$x_n = (a_i x_{n-i} + a_k x_{n-k}) \bmod m \qquad (4.4)$$

The linear congruential generator corresponds to k = 1 in the above recurrence. Taking m = 2 makes for a simpler implementation of the above generator but leads to a much shorter period and to major deficiencies (L'Ecuyer 1998). The following two sections discuss particular examples of multiple recursive generators.

4.2.3 Add-with-carry / Subract-with-borrow Generators

The add-with-carry (AWC) and subtract-with-borrow (SWB) generators are both examples of multiple recursive generators and were proposed by Marsaglia and Zaman (1991) (cited by Tezuka and L'Ecuyer 1992). The AWC and SWB methods can be viewed as variants of the additive and subtractive methods as discussed by Knuth (Knuth 1981). Both these generators combine efficiency and very long period (L'Ecuyer 1998).

The add-with-carry generator is based on the recursion (Tezuka and L'Ecuyer 1992):

$$x_n = (x_{n-s} + x_{n-r} + c_n) \bmod m \qquad (4.5)$$

$$c_{n+1} = I(x_{n-s} + x_{n-r} + c_n \geq m) \qquad (4.6)$$

where $r > s$ are positive integers called the *lags*, c_n is called the *carry* and *I* is the indicator function whose value is 1 if its argument is true and 0 otherwise. Again, a seed value x_0 must be selected to initiate the random number sequence. When *m* is large enough the normalisation of 4.2 can be used to find the random number on the unit interval. Although more generally, L successive values of x_n can be generated as follows:

$$u_n = \sum_{j=0}^{L-1} x_{Li+j} m^{j-L} \qquad (4.7)$$

The subtract-with-borrow generator is based on a similar recursion (Tezuka and L'Ecuyer 1992):

$$x_n = (x_{n-s} - x_{n-r} - c_n) \bmod m \qquad (4.8)$$

$$c_{n+1} = I(x_{n-s} - x_{n-r} - c_n < 0) \qquad (4.9)$$

As with the AWC generator, $r > s$ are positive integers called the *lags*, c_n is the *carry* and *I* is the indicator function whose value is 1 if its argument is true and 0 otherwise. Again a seed value must be selected to initiate the generator. The normalisations detailed above for the AWC are again used in this case to produce the U(0,1) variant.

The AWC and SWB generators both have extremely long periods, although not always maximal, in that the initially selected parameters to the algorithms do not result in the longest possible period. Additionally these generators have good distribution properties i.e. the random numbers produced follow the uniform distribution closely. However, these advantages can only be achieved when using a base other than a power of 2, which means the efficiency of the algorithm implementation is not optimal as the underlying architecture of the computer system is based on binary states. It has also been shown that these generators fail the spectral test in large dimensions (Tezuka and L'Ecuyer 1992) and this is not desirable as will be discussed in the Section 4.3.

4.2.4 Multiple-with-carry Generators

A generator of the same family as the AWC and SWB generators is the *Multiple with Carry* (MWC) generator. It has been shown that a power-of- 2 modulus can be used with the MWC generator while retaining a long period and potential for good properties (L'Ecuyer 1998, Marsaglia 1997).

In general the MWC generator is represented by:

$$x_n = (ax_{n-1} + c) \bmod m \qquad (4.10)$$

where an initial x and carry component c must be chosen along with an appropriate m. The new x_n is generated using the recursion of 4.10 and the new carry component c is the results of $(ax + c)/m$, which is the number of m's dropped when forming x.

The MWC sequence can be defined by the finite set S of ordered pairs (c, x), where c is the current carry component and x is the current output value (Marsaglia 1997). It is usual to express the ordered pair as ${}^{c}x$, thus giving the finite set S and the function f operating on it as:

$$S = \left\{ {}^{c}x : 0 \le c < a, 0 \le x < m \right\}, \qquad f({}^{c}x) = {}^{[ax+c/m]}(ax + c \bmod m)$$

It is practical to choose *a* and *m* so that $b = am - 1$ is prime (Couture and L'Ecuyer 1995, Goresky and Klapper 2003).

The more commonly used normalisation $u_n = \frac{x_n}{m}$ used in the generators above mean each output yields a multiple of *1/m*. To reduce the discretization error U_n may be constructed from several x_j's as in 3.7 which allows for smaller *m*, even *m=2*.

The MWC generator has properties of large period (but not maximal) and efficient computation (by using a base that is a power of 2). The uniformity of the random sequences has been shown to be poorer for this generator compared to those discussed previously (Couture and L'Ecuyer 1997). However, it has been shown that with slight modification to the MWC generator the maximal period can be achieved (Goresky and Klapper 2003). This class of generator will have a long life in the modelling world as its properties are understood more thoroughly.

4.2.5 Shift Register Generators

The shift register class of random number generators is split into two sub-classes, the linear feedback shift register (LSFR) and the generalised feedback shift register (GFSR). The name of this generator has its origin in the way that it is implemented, in that the bit representation of each random number is "shifted" left or right in the computer memory.

The linear feedback shift register (LFSR) is based on the recurrence:

$$x_n = (a_1 x_{n-1} + \ldots + a_k x_{n-k}) \bmod 2 \qquad (4.11)$$

where k > 1 is the order of the recurrence, $a_k = 1$ and $a_j \in \{0, 1\}$ for each j. The maximal period of $2^k - 1$ is achieved if and only if its characteristic polynomial

$$P(z) = \sum_{i=0}^{k} a_i z^{k-i}$$

(where $a_0 = -1$) is a primitive polynomial modulo 2. Again the normalisation of 4.7 is used to produce the U(1,0) random number.

The generalised feedback shift register (GFSR) class of generators are well known for being fast but can be complicated to initialise. This generator uses a recurrence of the form 4.11 where each bit of the output is a linear combination modulo 2 of the bits forming the states (L'Ecuyer 2000b). The R250 generator is an example of a GFSR. It has been shown to be fast and have a long period that is suitable for Monte Carlo simulations and has been found to be among one of the best generators (Janke 2002). Although it has been reported that the generator causes severe problems in some Monte Carlo applications (Janke 2002) mainly in the area of the simulations of the Ising cluster algorithm and the Metropolis update algorithm. Investigation into the problem found that correlations between triplets of random numbers was the cause of the problems and amendments have been made to produce a new more reliable generator denoted as R250/521 (Janke 2002).

An interesting variant of the GFSR is presented in (Hellekalek 1998) which is both fast and reliable if the parameters are chosen properly. This is the *twisted* GFSR that produces a sequence $(X_n)_{n\geq 0}$ of w-bit integers by the rule:

$$x_{n+p} = x_{n+q} \oplus x_n A \qquad n \geq 0$$

where (w, p, q, A) are the parameters of the tGSFR and A is a square matrix (of size w*w) of binary entries. On most computer systems an integer is represented by 4 bits of computer memory, thus w would be 4 and each integer generated would be 4 bits long.

The 'TT800' tGSFR presented by Matsumoto and Kurita (1992) has a period of 2^{800} and there is strong theoretical support, as well as convincing empirical evidence (Hellekalek 1998, L'Ecuyer 1998) for this random number generator. A C++ implementation of the Mersenne Twister generator by Matsumoto and Nishimura (Cooke 1998) is another example of a twisted GSFR. The LSFB, MRG and the LCG generators all possess long periods and are memory efficient when the parameters are chosen appropriately. Whereas the GFSR is less memory efficient, the Twisted GFSR is as fast as the GFSR while making optimal use of memory. However, in recent testing results by L'Ecuyer and Simard (2006) the twisted GFSR TT800 of Matsumoto and Kurita (1992) is shown to fail significantly more tests in the TestU01 suite (see Section 3.3) than the Mersenne Twister generator.

4.2.6 Combining Generators

Because random number generators are so important in a number of areas of computer simulation and modelling there has been considerable investigation into combining generators of both the same and different types in an attempt to find an even better random number generator. It has been found that combining generators can yield well structured random numbers with long periods that are computationally efficient. Two classes of combined generators have been well analysed in recent years, these being combined MRG's and combined LFGR's (L'Ecuyer 1998, 2000a).

4.2.7 Non-linear Generators

All of the above generators have been linear random number generators. However, these are not useful for cryptography purposes as the D-dimensional vectors of successive random numbers produced by the linear sequences is too regular. This makes the sequences of numbers easier to spot, and therefore the algorithm used to produce the sequence could be determined.

For these reasons nonlinear generators are required for cryptography. Although nonlinear generators also produce sequences of numbers they are more difficult to detect and the algorithms are almost impossible to determine simply from those sequences. Non-linear generators have been shown to produce better results than linear ones of equal period lengths (L'Ecuyer 1998) but they run at much slower speeds and this is prohibitive when implementing large Monte Carlo simulations. Therefore, for the purposes of Monte Carlo simulation, linear generators suffice and thus the discussion on nonlinear generators will not be extended here.

4.3 Testing Random Number Generators

As indicated above, a good random number generator must have a long period but this is not sufficient to conclude the generator is indeed good. The random number generator must have proven structure and statistical testing must be employed to provide this proof. There are two types of testing that should be considered when analysing a random number generator: theoretical tests and empirical tests. In general, theoretical testing examines the overall randomness of the sequences produced whereas empirical testing looks at the performance of the generator in a particular simulation problem.

Theoretical testing allows "figures of merit" to be produced for a generator which in turn can allow very good predictions of how a random number generator will perform in empirical tests. The most difficult and most important aspect of theoretical testing is the correlation analysis of a random number generator. Over twenty years of experience has shown that theoretical testing is important in producing support for random number generators (Dwyer Jr. and Williams 2003, Hellekalek 1998, Kao and Tang 1997, 1998, Kao and Wong 1998, Knuth 1981, L'Ecuyer 1998).

One of the most important theoretical tests is the spectral test. Coveyou and MacPherson (1967) originally proposed the spectral test as a way of looking at the correlation between consecutive random numbers. The spectral test is most commonly applied to linear

congruential and multiple recursive generators. Essentially the spectral test determines the maximum space between adjacent planes made up from points obtained from the D-dimensional vectors of consecutive random numbers that make up the lattice structure, otherwise known as hyperplanes. Each hyperplane is the random sequence produced from the initial parameters and seed of a generator. Two other theoretical tests commonly applied to generators are the Discrepancy Test and the Beyer Quotient (Hellekalek 1998, Knuth 1981, L'Ecuyer 1998). Both of these tests, as with the Spectral Test, are providing "figures of merit" in terms of the lattice structure of the generator.

The tests discussed above employ two statistical approaches that are commonly used for testing random data: the Chi-Square Test and the Kolmogorov-Smirnov Test. Both of these are testing the goodness-of-fit of the generator to the uniform distribution. Knuth (1981) provides a full theoretical explanation of these tests. These tests are not only used for the theoretical aspects of testing but are also used in empirical testing. Knuth (1981) discusses ten empirical tests that are used to investigate the randomness of the sequences produced by the random number generator and versions of these test are implemented in both the Diehard and TestU01 suites discussed below.

Each empirical test is a simulation and if selected properly will cover a whole class of simulation problems. It is important that the tests are carried out in a sensible manner. It seems obvious that any tests should be implemented in a way that reflects the simulation problem and that each test is carried out with the same parameters. If these rules are not followed nothing can be deduced from the results, as they do not reflect the simulation problem and the parameters being used in the problem.

There are a number of empirical tests that have already been designed and are used extensively (Entacher et. al. 2001, Kao and Tang 1998, Kao and Wong 1998, Knuth 1981, L'Ecuyer 1998, Liang and Whitlock 2001, Wegenkittl 2001). It may seem easy to design a test but what is really important is that each test designed constitutes a prototype of the simulation problem and can measure different properties of the random numbers.

One well established battery of empirical tests is that of George Marsaglia which is available on CD-ROM and from the web-server http://stat.fsu.edu/~geo/diehard.html. This collection of fifteen tests, called the Diehard tests, assesses the randomness of 32-bit integers obtained from a random number generator (Marsaglia 1997). Each test in this battery requires 2,867,200 32-bit integer values. It has been shown that the Diehard tests are particularly good at highlighting any problems with a random number generator (Carr 2003, Hellekalek 1998), and therefore can also be used to produce evidence of how good a particular generator is. A brief explanation of the tests is given in Table 4.1 below and full details can be found at http://stat.fsu.edu/~geo/diehard.html.

Test No.	**Description**
1	Birthday Spacings Test
2	Overlapping 5-permuationTest
3	Binary Rank Test for 31*31 matrices
4	Binary Rank Test for 32*32 matrices
5	Binary Rank Test for 6*8 matrices
6	Bitstream Test
7	OPSO (Overlapping-Pairs-Sparse-Occupancy), OQSO (Overlapping Quadruples-Sparse-Occupancy) and DNA Tests
8	Count-the-1's Test
9	Parking Lot Test
10	Minimum Distance Test
11	3D Spheres Test
12	Squeeze Test
13	Overlapping Sums Test
14	Runs test
15	CRAPS Test

Table 4. 1 – Diehard suite of tests

L'Ecuyer and Simard (2006) have also developed a software library, TestU01, for the empirical testing of random number generators. This library is the most comprehensive suite of testing utilities available to date and provides the user with the flexibility to chose particular tests to perform and select the desired parameters. The TestU01 software also provides source code for a wide range of generators and provides batteries of tests, much like the Diehard suit where tests and parameters are fixed. The three batteries of test provided for testing random numbers on the unit interval are Small Crush, Crush and Big Crush. These batteries provide a full range of classical testing and more stringent statistical testing with the complexity of the tests increasing to the most rigorous tests being performed in the Big Crush suit. L' Ecuyer and Simard (2006) present results of these tests as performed on a range of random number generators, thus providing a valuable reference when choosing a random number generator.

L'Ecuyer (1998) suggests that, for sensitive applications, the user should test the random number generator specifically for their particular problem, or should try random number generators from totally different classes and compare the results. This later technique has been used in finding a good random number generator for the Pore Architecture Model discussed in Chapter 6 and the results are presented in Section 4.6.

There are always new tests being devised for testing the randomness of generators (Liang and Whitlock 2001, Prohkova 2000, Wegenkittl 2001), but it should be noted that no amount of testing could prove that a random number generator is flawless; it can only improve our confidence in that particular generator.

4.4 Choosing a Good Random Number Generator

Making a choice of random number generator depends entirely on the application it will be used for. This thesis is concentrated on the area of computer modelling and simulation, thus the results and conclusions here should not be considered as universally applicable to other

problem domains such as cryptography. This is because the area of cryptography has a different set of requirements for an RNG and further safeguards need to be employed so that the generator cannot be 'learned' easily.

There are a number of considerations that should be made when choosing a random number generator, one of which being the speed of the generator. In general, most applications that use a random number generator require one that is relatively fast and in simulations where billions of random numbers are involved this can be a critical condition.

The portability of a generator is also important in that it can be easily implemented in a standard high level language and can produce the same sequence of random numbers on a number of different computers and compliers. Repeatability of a random number generator should also be considered, meaning that the sequence can be reproduced time and again. This is important for program verification and also for the re-running of simulations and models. It is normal to use a seed with a random number generator to invoke repeatable sequences. However, the choice of seed values must also be chosen carefully and with knowledge of the effects a badly chosen value may have on any simulations.

It is important to be aware that a RNG that has been written by a programmer without specialist knowledge of the area of random number generation should be scrutinised closely. Although it may seem easy to write a RNG algorithm that produces what looks like a random sequence it is somewhat more difficult to produce software that produces a statistically sound algorithm. A key point noted by Park and Miller is "strange and unpredictable is not necessarily random" (Park and Miller 1988).

4.5 Random Number Generators Included in Testing

Chapter 6 introduces the Pore Architecture Model that is used to simulate soil structures. This model uses a stochastic method to produce the soil structure matrix and thus the RNG within

the model framework is used heavily throughout any realisation of the model. The aim of the following sections is to find a good random number generator for this modelling environment.

A number of different classes of random number generators have been sourced to allow the investigation of these generators within the Pore Architecture Model. The BeOS library of random number generators (Cooke 1998) provides a good variety of generators that are easily added to Pore Architecture Model. Each of the generators will be analysed by both the Diehard tests (Marasglia 1997) for their statistical properties and within the Pore Architecture Model for evidence of suitability within this simulation environment. The generators below have been complied into a library by John Cooke who sourced the original code from other sources (see http://www.bebox.dircon.co.uk/rng.htm for full details). A brief description is supplied below. The BeOS library provides the following generators and the period of each generator can be found in Table 4.2:

i. LCG – this is the "minimal standard generator" by Park & Miller (1988)

ii. R250 – a fast shift-register generator

iii. ISAAC – a fast random number generator suitable for cryptographic purposes

iv. Mersenne Twister – a twisted GSFR generator

v. Mother – George Marsaglia's "The mother of all random number generators".

These RNG's cover a range of different classes of generators with varying period lengths. The following section shows the results of empirical testing carried using each of the generators.

Generator	Period
LCG	$2^{31} - 1$
R250	2^{250}
ISAAC	2^{8295}
Mersenne Twister	$2^{19937} - 1$
Mother	$\sim 2^{250}$

Table 4. 2 – RNG Periods

4.6 Results

4.6.1 Diehard Results

Each generator was tested against the fifteen Diehard tests and the results are shown in Table 4.3 below. These tests are developed to test both the independence and uniformity of the random umber sequences. These tests are considered stringent tests, as they seem to be more difficult to pass than the more classical tests of Knuth (1981).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
LCG			#		#	+	#	+	+						
R250				#		#		#							
ISAAC															
Mersenne															
Twister															
Mother															

Table 4. 3 – Diehard results, + is a partial fail, # is a clear fail, empty cell is a pass

As can be seen from the table above the ISACC, Mersenne Twister and Mother generators pass all 15 tests indicating these generators have good statistical properties and sound structure. The R250 generator fails three of tests and would therefore be less likely to be used in future. The LCG generator failed six tests, thus this generator would be less likely to be considered for future Monte Carlo modelling purposes.

4.6.2 Results with RNG's Embedded in a Simulation Environment

Having identified the good, bad and the ugly generators in terms of their statistical properties it is important to also test the generators in the particular modelling situation of interest. The area of interest here is stochastic simulation of soil structures using the Pore Architecture Model described in Chapter 6. Although some of the generators above were identified as being poor, all the generators have been tested with the Pore Architecture Model. This allows us to investigate if the poorer generators in terms of statistical properties also result in poorer models.

The Pore Architecture Model was used to produce a soil structure realisation with each of the random number generators described in Section 4.5. To implement consistent tests, the input data for each realisation of the model was the same, with the average porosity of the training images being 6.37%. To identify the effect a particular random number generator has on the output of the model, the porosities of each simulated structure can be compared to the average porosity of the training images. Table 4.4 below shows average simulated porosity obtained from 10 realisations of the model using each random number generator.

As shown in the Table 4.4 there is no significant difference ($P > 0.05$) between the modelled porosities using different random number generators, except from the realisations obtained using the R250 generator. This generator produce an average porosity of 32.82% over 10 realisations in comparison to the target porosity of 6.37%. This is is due to the divergence of this RNG from statistical randomness. It was stated in Section 4.2.5 that the R250 generator had been shown to cause severe problems in some Monte Carlo simulations (Janke 2002)

and that particular problem has been shown here again. The R250 generator would not be an appropriate generator for use with the Pore Architecture Model as this generator produces random sequences that significantly affect the simulation results.

RNG	Modelled Porosity (ave. of 10 simulations)
LCG	6.41%
ran3	6.33%
R250	32.82%
ISAAC	6.29%
Mersenne Twister	6.27%
Mother	6.40%

Table 4. 4 – Average porosity of simulated structures using each random number generator

The remaining generators have very little effect on the porosity of the simulated structures. Hence, any of these generators may be chosen for use in the modelling framework. The decision as to which one to choose would come down to factors such as period length, speed and ease of implementation. The results shown in Table 4.4 show how important it is not only to test the statistical properties of a random number generator but to also test the generator in the particular program of interest. It is not always the case that a generator with reasonable statistical testing results will be a useful generator when in use with a Monte Carlo algorithm.

With relation to the modelling framework described in the previous Chapter, the Mersenne Twister is selected as the random number generator to be used. This generator has the longest period of all the generators analysed above and passes the entire suit of Diehard tests. Therefore, with the large quantities of random numbers required in the modelling

framework this generator is deemed most appropriate. Furthermore, L'Ecuyer and Simard (2006) show that the Mersenne Twister also performs well in battery of tests within the TestU01, where the all tests in Small Crush were passed, and only two tests were failed in both the Crush and Big Crush tests. Due to the nature of the Pore Architecture Model, and the modelling framework in general, it is necessary to have confidence in the quality of the random numbers being produced.

4.7 Discussion

There are a wide variety of random number generators available for use in computer simulations, each of them having particular advantages and disadvantages. In general a random number generator should possess properties of long period, good structure, efficient computation algorithm and should have proven test results both theoretical and empirical. The different recursion rules detailed above for generating random numbers on the unit interval are widely used and there is extensive support for these generators (Kao and Tang 1998, Kao and Wong 1998, L'Ecuyer and Simard 2006, Wichmann and Hill 2006). It is highly recommended not to invent new recursion rules but use one of the well-know generators that have been tested, applied to applications and have been shown to be successful. Further subjective aspects such as ease of implementation, portability and speed of the generators should also be considered when choosing a generator for simulation purposes.

In terms of Monte Carlo simulations it is important not only to test the statistical soundness of the generator but also to test the generator within the program where the Monte Carlo algorithm is implemented. As was shown in the previous sections, the R250 generator was reasonably sound in terms of statistical properties obtained using the Diehard suite but when used within the Pore Architecture Model the results were significantly different from the expected values. As there are now a vast number of easily accessible empirical tests available that are sophisticated in the mathematical aspects of the testing of a random number generator, confidence in the standard generators can be gained and the quality of the random numbers produced can be shown to be sufficiently random for simulation purposes.

The Diehard suit of tests provides fifteen tests that have proved to be useful in indicating any deficiencies with a generator, however the package does have some limitations. Firstly, the sample sizes used in the tests are not very large and this results in less stringent tests. Secondly, Diehard requires binary sequences of 32-bit integers for the test procedures to process, however many random number generators produce numbers with only 31 bit accuracy. Finally, the Diehard package does not offer the user any flexibility to select particular tests or change the parameters of a test, thus restricting the effectiveness of the testing package.

However, the TestU01 software library developed by L'Ecuyer and Simard (2006) provides a more comprehensive testing package that does offer the user the flexibility to select particular tests and parameters thereof. TestU01 also provides batteries of test that are seen to be more stringent than the Diehard tests. L'Ecuyer and Simard (2006) present results of a range of random number generators tested against the three Crush test suits providing valuable supporting evidence for these generators.

The previous chapter described the modelling framework that has been developed during this research and the Random Numbers package was identified as a core part of the software architecture. The development of this integral part of the system was informed by the findings presented here, leading to the implementation of the Mersenne Twister algorithm for the generation of random numbers within the modelling framework. The Mersenne Twister is selected for its long period and proven results across a spectrum of testing (L'Ecuyer & Simard 2006). The object-oriented design of the underlying software architecture allows the instantiation of many random number generator objects within any simulation realisation, thus providing any single algorithm or process within the model framework with its own random number stream.

In conclusion, the random number generators used in modelling are a crucial part of the overall system. With a badly chosen or implemented algorithm the results of simulated events

can be badly skewed, thus it is imperative to be well informed of the common pitfalls of some generators. The development of the Random Numbers package into the modelling framework provides good random number generators that are easily distributed to other researchers using the .Net Framework. Furthermore, this package is fully extendable to include more algorithms such as those to produce random numbers from normal, Poisson or Binomial distributions.

Chapter 5 – Quantification of Soil Structure

5.1 Introduction

The physical structure of soil plays a critical role in the functioning of the soil system as a whole. The distribution of void space, or pore space, particularly at the micro scale influences the transport of oxygen and nutrients round the system and thus determines how different parts of the soil will be able to function. With the use of computed tomography (CT) technology it is now possible to obtain images of soil structure at the micro scale in three dimensions (3D). The digital images obtained from CT scanning of a structure result in a 3D grid of voxels, the volume element of CT scans. Each voxel is allocated a greyscale value, which represents the electron density of the scanned area. Image processing can then be applied to convert the greyscale values to binary values that represent either a pore or solid element of the structure. The analysis of the distribution of the pore and solid voxels allow the structure to be studied and characterised in detail. The soil samples available for use in this study are introduced in Section 5.2. The main aim of this chapter is to show how the physical structure of soil can be characterised using a number of metrics that quantify the way the solid structure and pore space is distributed and to relate these measures to the function of the soil.

A suite of metrics has been developed that characterises the structure using eight different methods, ranging from simple metrics such as porosity, to more complex metrics like the spectral dimension. The suite of metrics that have been employed in this project are discussed in Section 5.3 and it is shown that the quantification of structure in this way can give insight into how the structure may function from a functional perspective. The implementation of some of the metrics into the modelling framework is non-trivial and the issues relating to implementation are discussed in Section 5.4. The 3D soil structures are quantified using the metrics and the results are presented in Section 5.5. It is shown the metrics can characterise the different range of porosities and textures found the imaged

structures as the different values obtained for the measures can be related to functional properties of the structures.

5.2 Methodology

The datasets presented in this study have previously been used in an investigation by Feeney et. al. (2006), where the spatial and temporal structure of soil exposed to a range of root-microbe interactions was examined. In Feeney et. al (2006), these datasets are used to show that by quantifying the soil structure, the organisation and function of the soil can be identified. These datasets provide not only the structural data but also information on the bulk levels of biotic activity as detailed in Chapter 1. This allows the validation between the quantification of structure via the suite of metrics described in the following section and relation of these metric values to the biological activity within the soil sample.

All samples were x-ray scanned to produce the CT measurements for the soil structure. The full details of the image processing techniques used can be found in Feeney et. al (2006). The CT scanning of the soil samples results in a digital representation of the soil structure where each voxel has a resolution of 4.4 μm and is represented by a single grey-scale value that relates to the brightness of the voxel ranging from 0 (black) to 255 (white). Cleaning and segmentation of these images is require to determine the soil aggregate, pore space and surrounding background that was captured during the scanning process and thus allow analysis of the internal soil structure.

Feeney et. al. (2006) employed a series of processing steps prior to segmentation to allow accurate determination of the soil and void phases. The first step was to reduce the grey-scale range, which resulted in the brightness of very bright features being reduced. This was required as these bright features masked darker features in their vicinity and resulted in a loss of detail at the segmentation stage. Having reduced the grey-scale range, two smoothing steps were applied to the images using a Sigma filter, which reduces noise but maintains any fine structures present. Segmenting the smoothed images using an adaptive grey-scale

segmentation algorithm then produced binary images for analysis. This segmentation algorithm subtracts a low pass filter of the smoothed image from the smoothed image and segments the resultant image according to a defined threshold value. The low-pass filter updates the grey-scale value of each voxel with the average value of the surrounding voxels within a defined box size; in this case a box of size 255 x 255 was used. This adaptive grey-scale segmentation was carried out twice on the images, once with a low threshold value and once with a high threshold value to ensure as much soil aggregate was captured in the final image as possible. Further processing was carried out using binary morphological closing and filling operations to remove any gaps between features and narrow crack or pores were detected using a valley finding algorithm on the original image and added to the binary images. Finally, to distinguish the image background from the internal pores, the pore space was eroded slightly and the colour grey was assigned to these pores. Thus, the final image resulted in three distinct phases: solid matrix displayed in white, pores in grey and background in black. A central volume of the original soil samples was then selected in an attempt to avoid edge effects in the final structures and also remove the background so only the two phases of interest were present in the final images. Therefore, the images could be transformed into a binary format and thus allow the characterisation of each structure using the suite of metrics introduced in the following section.

At present CT technology cannot pick up biotic phases within the soil samples easily, and thus cannot provide a means to study the influence of soil structure on biotic activity. However, the sample sets being used in this study were exposed to different levels of biotic activity by using planted and non-planted soil cores. The soil samples will be referred to as sample set 1, 2, 3 and 4 and relate to the zero-day, control, bulk soil and rhizosphere sets examined in Feeney et. al. (2006). Each sample set contains a number of replicates taken from the soil cores exposed to different levels of biotic activity. Samples sets 1 and 2 were exposed to low levels of biotic activity, sample set 3 was medium activity and sample set 4 was high biotic activity. Sample set 1 contains soil samples taken at zero-day incubation and gives a base line for the structural genesis of the habitats. Sample set 2 had no planting in the samples, thus this data set relates to the reconstituted arable soil that was the initial

prescribed structure for all samples. Sample set 3 contains samples taken from the bulk soil surrounding a planted region within the soil cores after 30 days incubation and Sample set 4 is comprised of samples taken from the rhizosphere area of planted cores after 30 days incubation.

5.3 The Metrics

To investigate the influence of soil structure on transport processes and biological activity within the soil, the physical structure must be characterised. Both the pore volume and surface area have a major influence on the chemical, solute and gas transport in the soil microenvironment. Thus the size and shape of the pore space of a structure must be quantified. This must be done in a way that allows comparisons between different structures and gives insight into the effect structure has on the transport of nutrients and subsequently on the biological activity within the soil system. This section describes the suite of metrics that have been developed to this end.

The quantification metrics can be spilt into two types: those that characterise bulk properties, and those that characterise the scaling properties of elements of the structure. The metrics that characterise bulk properties are porosity (ρ), vugh porosity (ρ_v) surface connected porosity (ρ_{sc}), the fraction of pore space surface connected (γ) and the solid voxel distance distribution. The metrics that characterise the complexity of a structure are the fractal dimension, the spectral dimension, and the interior design metrics.

5.3.1 Porosity Measurements

5.3.1.1 Porosity

The porosity (ρ) of a structure indicates the fraction of the total structure that is void space (porous) and is one of the most common measures used in soil science.

$$\rho = v_p / v \qquad (5.1)$$

where v is the total number of voxels and v_p is the number of pore voxels contained within a structure.

5.3.1.2 Surface Connected Porosity

The surface connected porosity (ρ_{sc}) of a structure is defined as the fraction of the whole volume that is porous and is connected to the outer environment. This is a useful measure as it is the connected pore pathways that influence the transport of resources around the habitat space (Rappoldt and Crawford 2001).

$$\rho_{sc} = v_{sc} / v \qquad (5.3)$$

where v_{sc} number of surface connected voxels.

5.3.1.3 Vugh Porosity

The vugh porosity (ρ_v) is the fraction of the total volume that is pore space not connected to the outer environment (Horgan and Ball 2005). This metric is calculated in the following way:

$$\rho_v = \rho - \rho_{sc} = v_p - v_{sc} / v \qquad (5.2)$$

5.3.1.4 Fraction of Pore Space Surface Connected

The fraction of pore space surface connected (γ) indicates the fraction of porous volume that is connected to the surface and is computed as follows:

$$\gamma = v_{sc} / v_p \qquad (5.4)$$

5.3.1.5 Solid Voxel Distance Distribution

The distribution of shortest distances between solid voxels and their nearest surface connected pore voxels is a further metric used to quantify structures in this study. This metric results in a frequency distribution showing the number of solid voxels as a function of the shortest distance from a surface connected pore. This distribution mimics the oxygen distribution found when all the pores below the resolution of the structure voxels, i.e. all the solid voxels, are effectively water filled (Rappoldt and Crawford, 2001). Thus, this metric gives an indication of one of soils key functions – that of aeration.

The ρ_{sc}, ρ_v , γ and distance distribution metrics are not uniquely defined as they depend on the definition of connectedness in a cubic lattice. Therefore, a detection neighbourhood is employed within the computation algorithms for these metrics. A full description of the neighbourhoods and their influence on the resultant metric values is discussed in Section 4.4.

5.3.2 Pore Complexity Measurements

The metrics discussed in the previous section measure the bulk properties of the pore and solid phases at a specific measurement resolution. The metrics introduced in this Section measure the size but also indicate the scaling properties and complexity of the porous regions within the sample. The complexity of the pores play an important role in understanding the functioning of the soil microenvironment because the pore networks act as transport pathways and these determine how oxygen and nutrients can move through the system to microbial active sites. The fractal and spectral dimensions introduced in the following sections provide quantifiers that can be used to characterise the scaling properties of transport coefficients within the soil microenvironment (Crawford et. al. 1993, Crawford et. al. 1995).

A further set of metrics have been developed to quantify the internal properties of the pore networks, these are the 'interior design metrics'. These characterise the morphology of independent pore clusters within a structure. Understanding the internal properties of a pore cluster is important as two pores with the same volume may have very different shapes – one

may be long and thin and may maximise the transport of oxygen to the surrounding soil, whereas if the pore was spherical in shape, the surface to volume ratio is minimized and this will limit the availability of oxygen to the soil matrix in the immediate vicinity (Serra 1982).

5.3.2.1 Fractal Dimension *(d_f)*

There is much literature that shows that fractal geometry and related dimensions can be useful in approximating soil structure (Chapter 1). The fractal dimension, or more precisely, the mass fractal dimension of, in this case, the pore space has been shown to be a useful metric in understanding the static processes that occur in soil (Crawford, et. al 1993, Crawford, et. al 1995). Furthermore, the fractal dimension has also been related to the moisture-release curve and can be used as a method of relating structure to function (Crawford et. al, 1995).

In essence the fractal dimension (d_f) is a measure of the clustering properties of the pore space. Values of the dimension that are close to 3 indicate a structure that smoothly covers 3 dimensional space. Values less than 3 indicate a distribution that covers the space incompletely and results in pore-free regions on all scales. Furthermore, this clustering of pore space is repeated across all scales so that the patterns appear the same at whatever resolution the structure is viewed (Pachepsky, et. al., 2000). The method employed here uses the box counting method to estimate the fractal (strictly the Hausdorff) dimension (Baveye, Parlange and Stewart 1998).

A plot of the log of the number of boxes required to cover the pattern as a function of the log of the box size gives a straight line whose slope is the fractal dimension. Whilst the fractal dimension gives an estimate of the clustering of the space, it does not give a measure of the density of the distribution. Therefore, we could have structures with lots of small pores with the same fractal dimension as structures with a few larger pores. The intercept of the plotted line described above with the vertical axis gives a measure (called the lacunarity) of the density of the distribution and a pattern with the same fractal dimension but with smaller pores would have a larger intercept (Mandelbrot, 1982). Therefore, both the slope and the

height of the plot characterise the distribution. However, it usual to use only the fractal dimension, i.e. the slope, in any description of a structure.

5.3.2.2 Spectral Dimension (d_s)

The fractal dimension is not on its own sufficient to describe the heterogeneity of a structure for the purposes of understanding the dynamical process that take place within the structure. It has been shown that the spectral dimension can be used to describe the diffusion and convection processes that occur within the soil system (Anderson et. al. 1998, Crawford, et. al 1993). The spectral dimension is a measure of the tortuosity of the pore network and this can be used to relate properties of soil structure to the transport of oxygen through the pore networks.

The spectral dimension (d_s) is a dimensionless factor always greater than 1 that indicates the tortuosity or degree of complexity of the surface connected pore space (Anderson et al., 1998). The spectral dimension of a structure is obtained by conducting a set of random walks through the surface connected pore space of the structure, counting the number of distinct voxel sites visited ($S(n)$) at each step, n, of the random walk, where a distinct voxel site is a voxel that has not previously been visited. Therefore, the spectral dimension characterises the extent to which the pore structure impedes the progression of a random walker from its starting point. It is shown that a plot of $\log S(n)$ versus $\log n$ yields a line with slope $d_s / 2$. Higher values of d_s indicate a more continuous, less tortuous pore space than values closer to 1, which indicates a more complex and tortuous pore network.

The process of calculating this metric for any given structure is not trivial. To obtain the value of d_s firstly requires a number of random walks through the surface connected pore space to be performed. This in itself requires much processing to be carried out on the 3D voxel array to accurately identify distinct sites, visited sites and termination of the walk. Secondly, a straight line must be fitted through the plot of $\log S(n)$ versus $\log n$ for each random walk in order to obtain the gradient and thus d_s for a single walk. The d_s value for the structure as a whole is then calculated as the average value of d_s across all random walks.

5.3.2.3 Interior Design Metrics

The characteristics of the interior of the pore networks are also an important feature of the soil structure. Two structures may have the same bulk porosity, or as discussed above the same fractal dimension, but a structure that comprises of many small individual pore clusters will function differently, in terms of oxygen, water and nutrient transport, than a structure comprising of one large connected pore cluster.

To analyse pore clustering of the CT-imaged structures in more depth, an algorithm was developed to search the 3D grid of voxels and identify each individual pore cluster. The implementation of this algorithm is discussed in Section 5.4.2. The state of each pore cluster can be described in terms of volume, surface area, whether it is connected to the outer environment and its location within the 3D grid. The algorithm developed to compute these metrics all require the surface connected pores to be identified prior to execution. Thus, the detection neighbourhood within the algorithm implementation may affect these metric values. The metrics developed to analyse the interior design characteristics of a structure are:

- Total number of pore clusters
- Fraction of pore clusters that are surface connected
- Surface area / volume distribution of pore clusters

This set of metrics has been developed in an attempt to quantify the morphology of the internal pore networks. The area of mathematical morphology as applied to image analysis provides many techniques that comprise of complex geometrical transformations and measurements that can describe the complexity of the morphology of a 3D image (Serra, 1982). However, the interior design metrics developed here use more straightforward algorithms that result in effective metrics that provide details of the internal pore networks but which do not require the intensive computation and time requirements that are integral to the morphological quantification as described in detail by Serra (1982)

5.4 Implementation and Validation

5.4.1 Detection Neighbourhoods

All of the metrics described above, except the porosity measure and fractal dimension, depend on the definition of connectedness of the voxels within a cubic lattice. Therefore, a detection neighbourhood is employed within the algorithms used to compute the metrics. The detection neighbourhood determines the number of neighbours any target voxel has, thus defining the way in which voxel connections are made. The modelling framework provides the flexibility to define different neighbourhoods therefore allowing the sensitivity of the metric values to different detection neighbourhoods to be examined. For the purpose of computing the metrics each voxel is treated as a dimensionless cube with three detection neighbourhood schemes defined for use in the computation algorithms. These are faces (nearest 6 neighbours), edges (nearest 10 neighbours) and vertices (nearest 26 neighbours).

The faces scheme defines each voxel adjacent to the six faces of the target voxel as the only neighbours used in the connection of pores. The edges scheme defines the six voxels adjacent to the faces of the target voxel and the four voxels on the corner edges as neighbours. The vertices scheme defines the target voxel be in the centre of a 3*3*3 cube of voxels, thus using this scheme a target voxel has twenty-six neighbours. The neighbourhoods that define greater number of voxel connections are expected to yield higher values for the computed metrics. However, will the detection neighbourhoods make a significant difference in the resultant metric values?

5.4.2 Computation Algorithms

The development of the algorithms to compute the metrics described above is informed by the design of the underlying object model as discussed in Chapter 3. The algorithms developed here are all embedded within the Structure class implementation and are only accessed through the calling of methods of a Structure object. The computation of the various metrics requires the traversal of the 3D array of Voxel objects (a Structure class attribute) to

identify their state and thus allow the calculation of the metric values for the Structure object being analysed. The computational algorithms for each metric are implemented independently within the modelling framework to facilitate flexibility. This results in a traversal of the Voxel array for each metric computation. While a single traversal of the Voxel array could have been used to implement many metric calculations in one pass this would compromise the flexibility of the development, maintenance and application of the metric calculations. While this is inefficient in terms of execution time, it allows more flexibility within the model framework and to the end user application.

The 3D array of Voxel objects is a physical representation of the 3D soil structure; within this there are logical relationships between Voxel objects, in particular the connectivity of pore voxels to neighbouring pore voxels. This connectivity relationship can be considered as a graph or network of pore voxels. This logical relationship is defined by pore voxels being neighbours within the voxel array as defined by the detection neighbourhoods discussed in Section 5.4.1. Central to many of the algorithms is a traversal of the logical connected pore network, and standard network searching strategies are employed in the form of either a depth first search or a width first search (Knuth, 1973). The benefits of using these search strategies to compute the different metric values are discussed in the remainder of this Section.

The algorithm implemented to calculate the surface connected porosity along with the other porosity metrics, uses a depth-first search of the logical pore network within the 3D Voxel array. The pseudo code for this algorithm implementation is shown in Figure 5.1. This search strategy is used to process each voxel within the structure and detect if it is a pore voxel and is connected to the surface This is implemented using a Stack object, which uses a last-in-first-out queuing system to store the Voxel objects that still require to be processed. The algorithm shown in Figure 5.1 results in pore voxels and their connectivity being identified from the outer layers of the structure moving towards the centre, i.e. each connected pore network within the structure is identified in a depth first fashion. Because each Voxel object maintains its location within the 3D array, the pore Voxels can be pushed onto the Stack

object for processing in any order and also provides an efficient way of detecting neighbouring Voxel states. The implementation of the algorithm in this way results in an efficient traversal of the connected pore network of a structure.

```
Detect Surface Connected Pores

Initialise stack with all pore voxels from top and bottom layers of the structure

While stack is not empty
{
   Get next target voxel from stack
   If target voxel is not connected
   {
      Set SurfaceConnected property to TRUE
      Get all neighbours of target voxel
      Any neighbour that is a pore is pushed onto stack
   }
}

Count total pore voxels
Count total surface connected voxels

Porosity = total pore voxels / total voxels
Surface Connected Porosity  = total surface connected voxels / total pore voxels
% pore space surface connected   = Surface Connected Porosity / Porosity
```

Figure 5. 1 – Algorithm to detect the surface connected pore voxels

The algorithm developed to find the distance distribution of solid voxels to the nearest surface connected pore voxel uses a width-first search of the 3D Voxel array. Prior to computing this metric, the surface connected pore voxels must be identified as described above. The relationships between the surface connected pore voxels and the solid voxels is again defined using the detection neighbourhoods as described previously. The pseudo code for this algorithm implementation is shown in Figure 5.2. This search strategy results in solid voxels distances being identified in a bubbling manner working out from the connected pore voxels. This is implemented using a Queue object that uses a first-in-first-out strategy to store the Voxel objects that are waiting to be processed. Each voxel object is processed to identify and set its distance property, thus the distance distributions can be computed by counting the total number of voxels over each distance class.

```
Find voxel distances

If surface connected pores have not been computed
Compute surface connected pore voxels

Initialise queue with all surface connected pore voxels

While queue is not empty
{
   Get next target voxel from queue
   Get all neighbours of target voxel

   For each neighbour voxel
   {
      If neighbour is not connected
      {
         If nhb.DistanceToConnectedPore = 0 OR nhb.DistanceToConnectedPore > targetVoxel.DistanceToConnectedPore + 1
         {
            Set nbh.DistanceToConnectedPore = targetVoxel.DistanceToConnectedPore +1

            Add neighbour voxel to queue
         }
      }
   }
}
```

Figure 5. 2 – Algorithm to find the distance to nearest surface connected pore for all voxels

The design and implementation of the interior design metrics is more complex because the independent pore clusters within the 3D structure must be identified and measured in terms of volume and surface area. However, the object oriented design of the modelling framework allowed the inclusion of two classes to implement these metric calculations. The design of the Cluster Finder and Pore Cluster classes can be found in Appendix I. The Cluster Finder class provides the process of identifying all the independent clusters and their measurements while the Pore Cluster class provides a data structure to store and carry out analysis on the clusters. The algorithm to process and calculate the pore cluster metrics is shown in Figure 5.3. This algorithm is embedded within the Cluster Finder class. However, the functionality contained within these classes is only accessible through a method call to the Structure class and it is within the Structure class that the analysis of the pore clusters is carried out and reported on.

The algorithm identifies the first pore voxel that does not have a cluster number assigned to it. A width first search of the connected pore voxels is then carried out to identify all pore voxels belonging to this cluster and the volume and surface area is calculated. This is implemented

in a similar manner to the algorithm to calculate the solid voxel distances using a Queue object. The use of a width first search strategy in this computation algorithm is a more efficient search mechanism than using a depth first search within the context of the metrics being computed.

Having identified an independent pore cluster, a new Pore Cluster object is created that encapsulates the state of the cluster as shown by the attributes of the class (see Appendix I). This process is repeated until all pore voxels have an assigned cluster number. The Pore Cluster objects are sorted by volume and then by surface area within each volume class and stored in an array of arrays to allow analysis of the pore cluster properties of a structure via the interior design metrics described in Section 5.3.2.3.

5.4.3 Validation of Computation Algorithms

To systematically test the implementation of the quantification metrics within the modelling framework, a set of structures with known properties were constructed. These included high, medium and low porosity structures as well as completely pore and completely solid structures. More importantly, structures with known features such as loops in pore networks, complicated pore connectivity and clusters with know positions, surface areas and volumes were constructed to fully validate the implementation of the various metrics. CT-imaged datasets that have been analysed by other researchers were also used for verification.

```
Identify Clusters Algorithm

 cNo = 1

 for each voxel in Structure
 {

  if voxel is pore AND voxel.ClusterNo = 0      //new pore cluster found
  {
   Set voxel.ClusterNo = cNo

   set cluterVol = clusterArea = 0

   add voxel to queue

   while queue is not empty                     //start process of finding all pore
   {                                            //voxels belonging to cluster
    Get target next voxel from queue
    add 1 to clusterVol

    Get all neighbours of target voxel

    for each neighbour voxel
    {
     if neighbour is pore AND neighbour.ClusterNo = 0
     {
      neighbour.ClusterNo = cNo
      add neighbour to queue
     }
     else if neighbour is solid
     {
      add 1 to clusterArea
     }

    }                       // all voxels in cluster are identified

    create new Pore Cluster object with attributes set to:
                        neCluster.Volume = clusterVol,
                        newCluster.SurfaceArea = clusterArea,
                        newCluster.IsCOnnected = voxel.IsConnected,

    add 1 to cNo
   }

  }

 }
```

Figure 5. 3 – Algorithm to identify and quantify the independent pore clusters

5.5 Results

The metrics described in Section 5.3 are calculated for each of the replicates within each of the sample sets. For ease of presentation and explanation, the average metrics values and standard errors for each sample set are shown and discussed throughout the following sections.

5.5.1 Porosity (ρ)

The porosity of each replicate within each sample set was calculated by scanning the 3D grid of voxels and counting the number of pore and solid voxels that existed within the individual structure. The average porosity for each sample set is shown in Table 5.1. The sets ranged from 5.1% (Sample 1) – 15.1% (Sample 4) porosity. The standard error of Sample 1 is 0.6 indicating that the replicates do not vary greatly around the porosity of 5.1%, whereas Sample 4 has a standard error of 5.4 demonstrating that the individual replicates have porosity values ranging from 5.5% - 24.6% within the sample set.

5.5.2 Surface Connected Porosity (ρ_{sc})

The surface connected porosity (ρ_{sc}) is the first metric introduced that uses the detection neighbourhoods as described in Section 5.4.1. The ρ_{sc} metric values calculated using each of the three detection neighbourhoods are shown in Table 5.1. It can be seen that there is no significant difference ($P > 0.05$) in the resultant metrics when using the faces, edges or vertices neighbourhood schemes. However, the vertices scheme always yields in the highest surface connected porosity for any individual structure or sample set. This finding is expected due to the larger number of neighbours used in connecting pore voxels when this scheme is used. The ρ_{sc} metric for Sample 1 using the vertices neighbourhood is 2.9%, Sample 2 is 4.0%, Sample 3 is 5.8% and Sample 4 is 13.1%.

Metric	Porosity	Surface Connected Porosity			Vugh Porosity			%Pore Space Surface Connected		
Detection Nbh		Faces	Edges	Vertices	Faces	Edges	Vertices	Faces	Edges	Vertices
Sample 1 - Zero-day	5.1 (0.6)	1.6 (0.6)	1.6 (0.6)	2.9 (1.0)	3.5 (0.7)	3.5 (0.7)	2.2 (0.3)	28.6 (8.6)	28.9 (8.5)	50.1 (12.4)
Sample 2 - Control	5.7 (1.0)	3.7 (0.9)	3.7 (0.9)	4.0 (1.0)	2.0 (0.3)	2.0 (0.3)	1.7 (0.2)	64.3 (4.0)	64.4 (4.0)	67.4 (5.7)
Sample 3 - Bulk soil	8.1 (2.4)	5.4 (1.9)	5.4 (1.9)	5.8 (2.0)	2.6 (0.5)	2.6 (0.5)	2.3 (0.4)	60.6 (10.0)	60.7 (10.0)	64.9 (9.1)
Sample 4 - Rhizosphere	15.1 (5.4)	12.8 (4.1)	12.9 (4.1)	13.1 (4.0)	2.3 (0.6)	2.2 (0.6)	2.1 (0.6)	71.6 (13.2)	72.0 (13.2)	74.1 (12.2)

Table 5. 1– Porosity metrics obtained from the CT imaged soil structure data sets; these are the average values obtained from the replicates within each sample set, standard error shown in brackets. Results for each of the detection neighbourhoods are shown.

5.5.3 Vugh Porosity (ρ_v)

The vugh porosity is calculated for each of the individual structures using the three different neighbourhoods. The average results for each sample set can be found in Table 5.1. The detection neighbourhoods again have no significant effect ($P > 0.05$) on the metric values. The ρ_v values calculated, using the vertices neighbourhood scheme, for Sample 1 is 2.2%, Sample 2 is 1.7%, Sample 3 is 2.3% and Sample 4 is 2.1%.

The vugh porosity across all structures is shown to be similar and is usually the smallest element of the total porosity. However, in some low porosity samples the volume of pore space unconnected to the outer environment does exceed the surface connected pore space. This may be due to artefacts introduced in the image processing procedures where small pore elements have been falsely identified as soil matrix. However, it is also possible that the resolution at which the samples have been imaged (4.4 μm) has resulted in fine pores not being detected and therefore this anomaly occurs.

5.5.4 Fraction of Pore Space Surface Connected (γ)

This metric is again calculated for each structure using the three detection neighbourhood schemes and the average values of all replicates in each sample set are shown in Table 5.1. This metric is shown to be more sensitive to the detection neighbourhoods than the surface connected porosity or vugh porosity metrics. Yet, there are no significant differences ($P > 0.05$) found in the metric values obtained using the different neighbourhood schemes. However, as the porosity of the samples increase, the difference in the metrics values obtained using the faces and vertices schemes decreases, indicating that metric values calculated using detection neighbourhoods are more sensitive when dealing with low porosity samples due to the more fragmented nature of pore voxels. The fraction of pores space surface connected (γ) for Sample 1 is 50.1%, using the vertices neighbourhood with Sample 2 yielding a value of 67.4%, Sample 3 is 64.9% and Sample 4 has a value of 74.1%.

5.5.5 Solid Voxel Distance Distribution

For each replicate the frequency distribution of the distance to nearest connected pore for each solid voxel was calculated using each of the detection neighbourhoods. As expected the distributions obtained using the faces scheme resulted in distributions with a greater spread than those obtained using the vertices scheme (Fig. 5.4). It was found that high porosity structures resulted in maximum distances to nearest connected pore of approximately half the maximum distances found in lower porosity structures (Fig. 5.5). This is a consequence of the higher surface connected porosities of these structures.

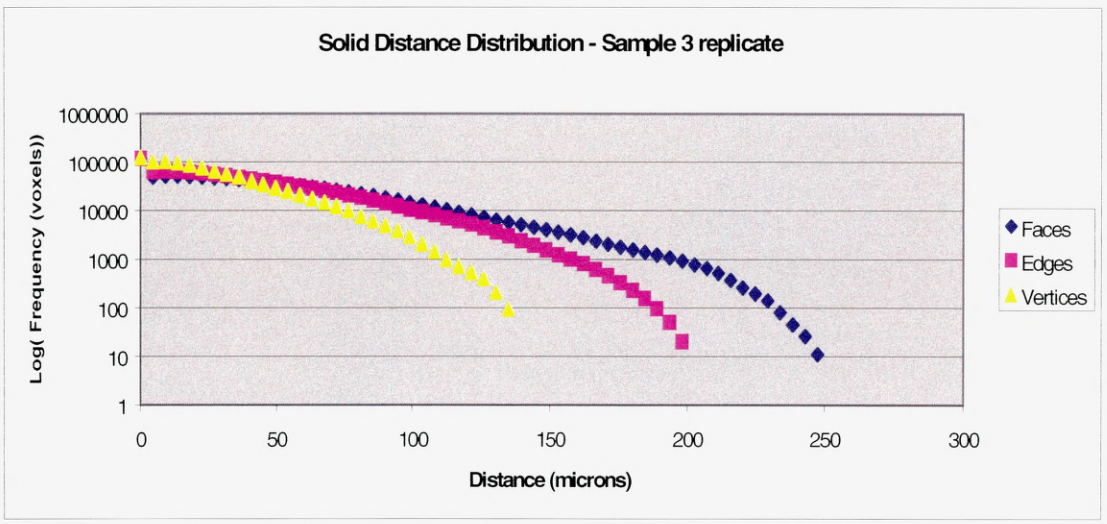


Figure 5. 4 - Solid distance distribution showing the effect the detection neighbourhood has on the spread of voxel distances

5.5.6 Fractal Dimension (d_f)

The average fractal dimension values obtained for each sample set are shown in Table 4.2. Sample set 1 yields a fractal dimension of 2.75 (standard error of 0.0.3). Sample set 2 has a value of 2.93 (s.e. 0.04), sample set 3 is 2.84 (s.e 0.05) and sample set 4 is 2.86 (s.e 0.56). The greater variation of the individual structure within Sample set 1 is again shown by the higher standard error value in comparison to the other sample sets. It is shown that Sample set 2 yields the highest fractal dimension indicating that these structures are the most

heterogeneous, while Sample set 1 is least heterogeneous. This indicates that pore structure is more space filling over all scales in the samples with higher fractal dimensions.

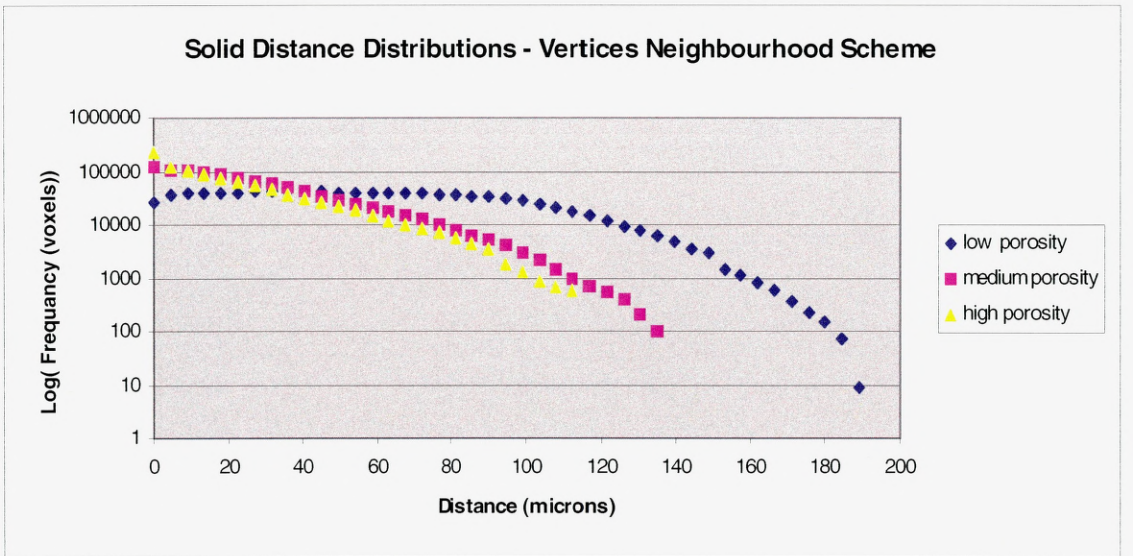


Figure 5. 5 - Solid distance distributions for low, medium and high porosity structures using the vertices detection neighbourhood.

5.5.7 Spectral Dimension (d_s)

The spectral dimension is a computationally intensive metric to calculate due to the number of random walks conducted within the surface connected pore space. The more walks carried out the more accurate metric value obtained and in this case 1000 random walks with a maximum of 60000 steps was initiated through each CT-imaged structure. It can be seen from Table 5.2 that Sample set 1 resulted in an average spectral dimension of 1.42, Sample set 2 is 1.47, Sample set 3 yielded a value of 1.54 and Sample set 4 resulted in 1.52. These data suggest that Sample set 1 (low porosity) has the most tortuous connected porous networks and Sample sets 3 and 4 (high porosity) have the least tortuous or complex connected pore networks.

Metric	Spectral Dimension	Fractal Dimension	Total Pore Clusters			%Clusters Surface Connected		
Detection Nbh			Faces	Edges	Vertices	Faces	Edges	Vertices
Sample 1 - Zero-day	1.42 (0.08)	2.86 (0.56)	920 (67.9)	848 (63.0)	757 (58.1)	6.6 (0.006)	7.0 (0.006)	3.1 (0.004)
Sample 2 - Control	1.47 (0.07)	2.84 (0.05)	872 (58.3)	832 (51.3)	745 (50.9)	5.1 (0.006)	6.1 (0.005)	2.8 (0.003)
Sample 3 - Bulk soil	1.53 (0.04)	2.93 (0.04)	839 (40.4)	786 (44.1)	668 (37.3)	2.9 (0.003)	3.0 (0.003)	3.6 (0.003)
Sample 4 - Rhizosphere	1.52 (0.07)	2.75 (0.03)	1039 (282.4)	968 (245.5)	788 (170.6)	3.5 (0.003)	3.5 (0.002)	4.1 (0.003)

Table 5. 2 – Pore complexity metrics; values shown are average values of replicates within each sample set, standard error shown in brackets.

5.5.8 Interior Design Metrics

The characterisation of the pore clusters contained in each CT-imaged sample via the interior design metrics was carried out and the results are presented in Table 5.2.

5.5.8.1 Total Pore Clusters

It can be seen that the metric values obtained for the total number of pore clusters is affected significantly ($P < 0.05$) by the detection neighbourhood used, with the faces neighbourhood resulting the greatest number of pore clusters and the vertices neighbourhood yields the lowest metric values. This is to be expected as the vertices neighbourhood provides more logical relationships, i.e. connections to be made between the pore voxels. In general the low porosity samples had more pore clusters than the high porosity samples. This indicates that the total pore volume in high porosity samples is more continuously connected than in low porosity structures.

5.5.8.2 Fraction of Pore Clusters Surface Connected

The fraction of pore clusters surface connected metric shows that detection neighbourhood used has a significant effect ($P < 0.05$) on the resultant values across all sample sets. The low porosity sample sets (sample sets 1 and 2) show a decrease in the fraction of clusters

surface connected as the detection neighbourhood increases in size. The resultant metric values for the low porosity samples are shown to decrease as the detection neighbourhood increased in size indicating that individual pore clusters are connected via small pore pathways that are formed as logical relationships using the larger detection neighbourhoods. However, the high porosity structures (sample sets 3 and 4) show an increase in the metric value as the detection neighbourhood size is increased. This indicates that pore clusters that were previously unconnected to the surface have been provided with a pore voxel connection to the surface by the increased size of the detection neighbourhood.

5.5.8.3 Surface Area / Volume Distribution

An example of the surface area to volume distribution of pore clusters for a high, medium and low porosity sample is shown in Figure 5.6. It can be seen that the low porosity sample yields more data points than the high porosity sample, indicating the greater number of pore clusters. It should be noted that each sample shows a single pore with a high volume and high surface area, the raw data shows that in every sample one large pore is always connected to the outer environment. This suggests the majority of the surface connected porosity is made up of one large correlated pore region.

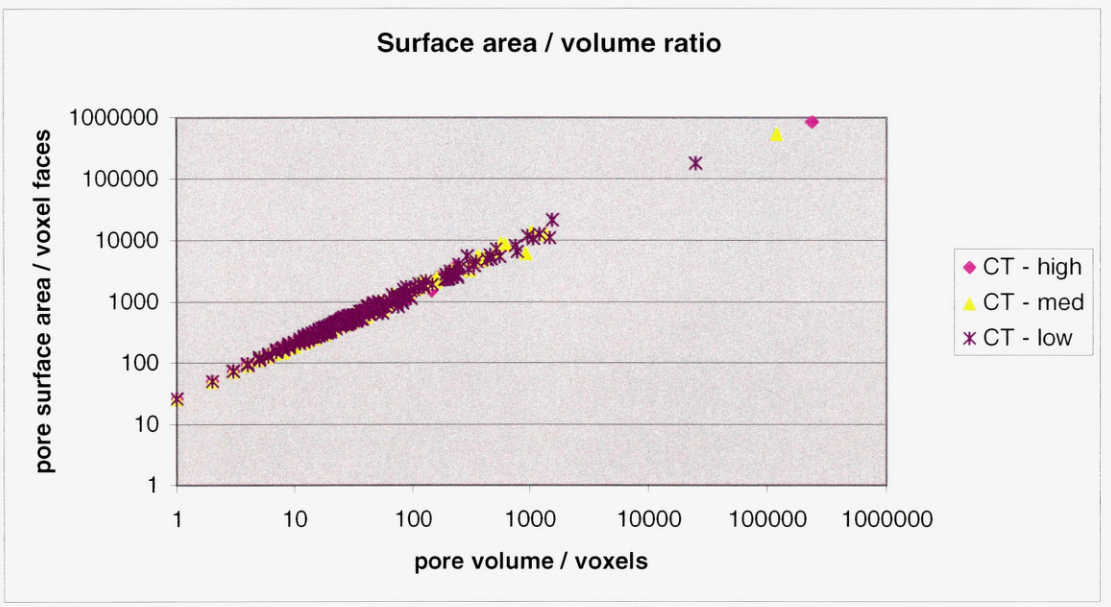


Figure 5.6. Pore cluster surface area / volume ratio

5.6 Discussion

The structure of soil influences the transport processes that take place within the soil system. In particular, the size and connectedness of the pore space plays a major role in the flow of gases and fluids around the system with more continuous pore networks facilitating gas diffusion into the soil matrix. These transport processes are fundamental to the biological activity that can be sustained within the microenvironment. However, there is no comprehensive theoretical framework with which the physical structure can be quantified and directly related to biological function. The metrics presented in this Chapter aim to provide a suite of metrics that quantify the different aspects of soil structure and that the metric values obtained can be related to the soil function.

The soil samples analysed using these metrics were shown to span a range of porosities and textures. The analysis shows that in general, the higher porosity structures (sample sets 3 and 4) had more surface connected pathways than lower porosity sets (Table 1). This would be expected as sample sets 3 and 4 were exposed to higher levels biotic activity as these samples were taken from planted cores where plant roots and associated microbial activity would result in more porous pathways being created. Whereas sample sets 1 and 2 were taken from unplanted cores where biotic activity would be less significant due to the absence of plant roots.

The total porosity for a high, medium and low porosity sample can be seen visually in Figure 5.7 (a) along with the porosity value (i) and fractal dimension of the pore space (ii). It can be seen that the high porosity structure yields the lowest fractal dimension and the low porosity structure yields the highest fractal dimension. This indicates that the high porosity structure, which is almost 25% pore space, is more heterogeneous in terms of the distribution of the pore volume within the structure as shown by the fractal dimension of 2.74. Whereas a low porosity structure, which has a porosity of only 4.85%, has a fractal dimension of 2.99 indicating the pore volume more smoothly covers the 3D space of the structure, i.e. this low porosity structure is less heterogeneous compared to the high and medium porosity structures.

The porosity and fractal dimension metrics, if used alone, are of limited use in relating the physical structure of soil to biological functioning because these metrics describe properties of the total air filled porosity of a structure. However, it is the pore networks that have access to the outer environment that govern the transport of oxygen, water and nutrients around the soil system. Therefore, the use of the surface connected porosity metric and the spectral dimension of the surface connected pore space are metrics that enable a structure to be quantified and the resultant metric values can be related to the transport properties of the structure. The total surface connected porosity for the high, medium and low porosity structures can be seen visually in Figure 5.7 (b) along with the surface connected porosity value (i), the fraction of pore space surface connected (ii) and the spectral dimension of the surface connected pore space (iii). It is clearly seen that the total surface connected pore volume, which facilitates the transport of oxygen etc. around the soil system is significantly different across the structures. Thus, the transport properties of these structures will be different and can be defined by the connectedness and tortuosity of the pore volume as quantified using the surface connected porosity and the spectral dimension metrics respectively. It is evident from the spectral dimension metric values that the low porosity structure provides the most tortuous surface connected pore networks as indicated by the value of the spectral dimension which is 1.34, whereas the medium and high porosity samples are similar in terms of tortuosity with values of 1.67 and 1.63 respectively. However, the medium porosity structure is shown to have the least tortuous pore network even although this structure is almost half as porous as the high porosity structure. This indicates that the medium porosity structure is comprised of smaller more complex pore pathways.

In an attempt to describe the geometry and topology of the internal pore clustering, the interior design metrics have been employed. These metrics are shown to provide further detail on the correlation of the pore space within a given structure. It can be seen in Figure 5.7 (c) that each sample contains a single correlated pore cluster that is connected to the outer environment and that this independent pore cluster makes up the majority of the surface connected porosity. This situation is found across all the samples studied and shows that

although the soil microenvironment is complex, the important features that govern transport around the system can be identified and quantified in simple terms.

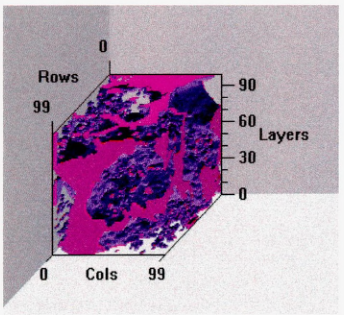
The fraction of surface connected pore space shows that lower porosity structures, as expected, have far less of the total pore volume connected to the outer environment than high porosity structures. As it is the surface connected pore space that governs the transport of oxygen through the soil structure the samples with higher surface connected porosity will be more likely to maintain a more oxygen rich environment and hence be capable to supporting higher levels of microbial activity. The distribution of solid voxels in relation to surface connected pores also relates to the distribution of oxygen in the soil matrix. Thus, this metric is also useful in indicating level of aeration within a structure. A visual representation of the solid voxel distance distribution for low, medium and high porosity samples can be found in Figure 5.7 (d). It is clearly seen that the high porosity structure can enable the diffusion of oxygen into the surrounding soil matrix more effectively than the medium and low porosity structures. The high porosity structure shows fewer red and yellow areas, which indicate longest distances to surface connected pores, in comparison to the medium and low porosity structures which show larger volumes of the structures having longer distances to nearest surface connected pores.

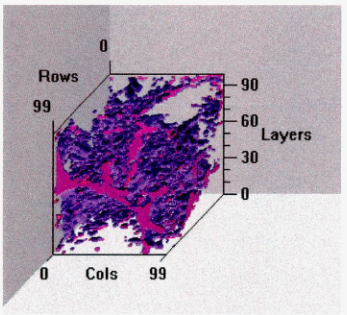
The use of a detection neighbourhood for computing the various porosity measures was examined and, as expected, the larger neighbourhoods resulted in higher metric values than when a smaller neighbourhood was used. This outcome is due to the larger neighbourhoods providing more logical connections between voxels. It can be noted from Table 1 that the different detection neighbourhood used for computing the ρ_{sc} and ρ_v metrics on the CT-imaged structures showed no significant difference in the resultant metric values. However, the metric values obtained for the fraction of pore space surface connected (γ) using each of the detection neighbourhoods showed more variation across sample sets (Table 5.1). However, only the low porosity structures (Sample set 1) show a significant difference between the faces and vertices neighbourhoods. This indicates that the pore voxels within the

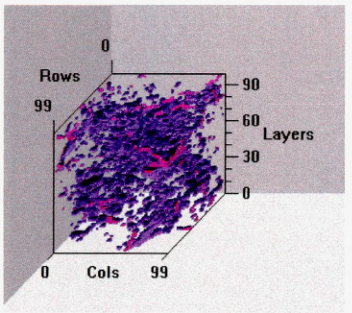
low porosity structures are more sparsely distributed than the higher porosity structures where the detection neighbourhood had negligible effect.

In conclusion, the suite of metrics introduced in this Chapter provides a quantitative framework with which to quantify soil structures. In addition, it has been shown that by using the range of quantification metrics as described in Section 5.3, the resultant metric values that describe the physical structure should be able to be related to the function of soil in terms of oxygen and solute transport around the microenvironment.

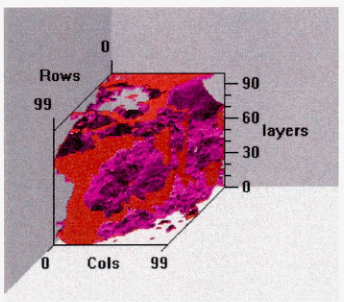
(a) (i) ρ	24.63%	13.6%	4.85%
(ii) d_f	2.74	2.88	2.99

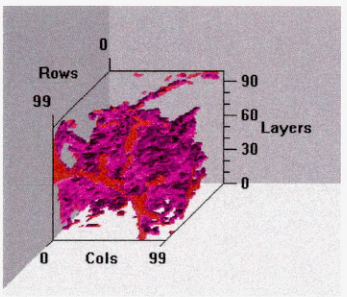


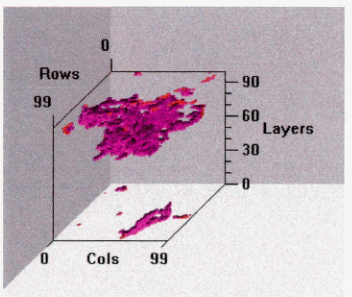




(b) (i) ρ_{sc}	23.63%	12.35%	2.67%
(ii) γ	95.94%	90.81%	54.98%
(iii) d_s	1.63	1.67	1.34

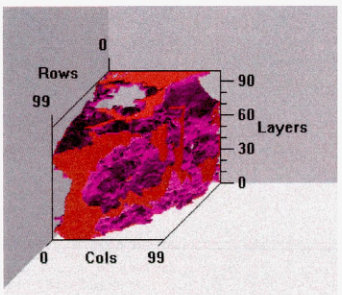


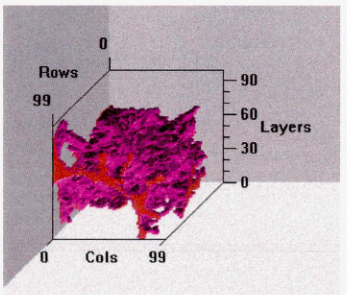


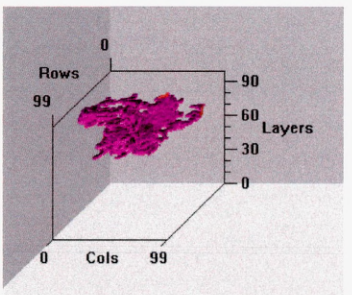


(c) Single surface connected pore cluster

(i) Cluster volume	23.55%	11.91%	2.49%







(d) Solid Voxel Distance Distributions

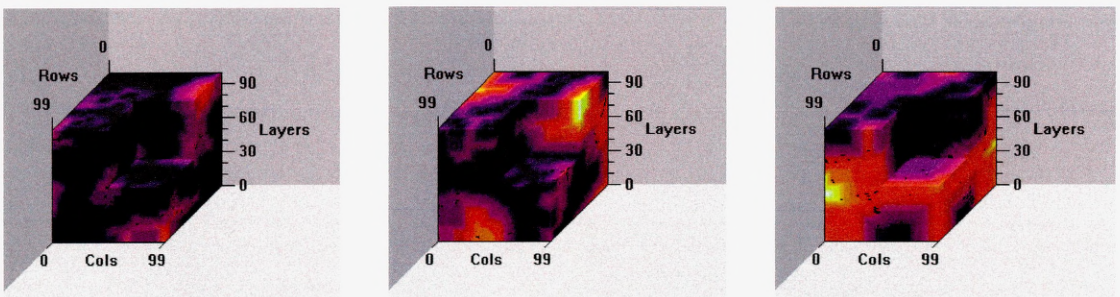


Figure 5.7 - Visual representation of pore scale properties for a high, medium and low porosity CT structure.

Chapter 6 – The Pore Architecture Model

6.1 The Pore Architecture Model

This Chapter introduces a model that is shown to simulate the key features of 3D soil structure at the micro scale using data from 2D thin sections of soil. It is important to be able to model soil structure in three dimensions as CT scanning of soil is still, at this time, unable to pick up all phases of the soil including the biotic phases. Most data relating to soil structure is available from biological thin sections, which can only provide structural information in two dimensions. Therefore, a model of soil structure will enable the investigation of the influence of structure on the biological functioning of the soil microenvironment. A model that provides good simulated structures also provides a means of investigating many different microenvironments.

The Pore Architecture Model has been developed and implemented using the same object-oriented approach as described in Chapter 3 where each mathematical component of the model is designed as a class with properties and methods describing the state and behaviour that an object of that class can exhibit. The underlying object model can be found in Appendix I. This design methodology has proved effective in developing the underlying mathematical model as the class design has facilitated the inclusion and testing of different algorithms that carry out specific parts of the simulation process. This has allowed the comparison between the algorithms and the identification of the optimum algorithm, which is presented in this Chapter.

The method used to simulate the 3D microstructure of soil in this study is an extension of the 2D model described by Wu et al. (2004). The model presented in this Chapter is an implementation of the model presented by Wu et. al (2006). The model algorithm is shown as a flowchart in Figure 6.1. The following section describes how training images are parameterised. Section 6.1.2 details how the parameterised data is used to build a three-dimensional matrix representing the soil structure and Section 6.1.3 outlines the methods

employed to optimise the clustering properties of the 3D matrix. Section 6.2 discusses the validation of the model and the results are presented in Section 6.3 with a discussion of the model presented in Section 6.4.

6.1.1 Model Parameterisation

The 2D training images used to generate a 3D structure simulation must be representative of the structure as a whole. It is known that soil structure can be heterogeneous and anisotropic in nature and therefore the statistical properties of the spatial structure maybe different in each orthogonal plane. If the parameterisation of the model is obtained from only one plane, the resulting simulated structure may not accurately reflect the distribution of pore and solid regions in three dimensions. Therefore, this model uses three orthogonal thin sections, representing a vertical, horizontal and perpendicular slice from a soil sample in an attempt to capture a more complete set of spatial data. The known 2D information in this study comprises three orthogonal slices extracted from the CT images introduced in the pervious Chapter. These data are used to parameterise the model via the calculation of conditional probabilities.

Each training image is used to derive a set of conditional transition probabilities dependant on a set of local neighbours. These probabilities are calculated by successively applying the six-cell neighbourhood shown in Figure 6.2 to each pixel element in the 2D training image. The cells located at point (i, j) and (i, j+1) in Figure 6.2 are the target voxels. Each voxel in the training image is considered to be at the (i, j) position of the six-cell neighbourhood and the combination of voxel states in relation to, and including the target voxels are recorded. To calculate the transition probabilities, a count of the number of occurrences of a particular neighbourhood state combination is taken. This number is then divided by the total number of neighbourhoods in the training image to result in the conditional transition probability for that neighbourhood state combination. Transition probabilities are also calculated for two-cell neighbourhoods to allow initiation of a structure simulation and boundary conditions to be handled. This method of model parameterisation is the same method used by Wu (2004) in

the modelling of 2D structures. Additionally, conditional probabilities are calculated in the same manner for a nine-cell neighbourhood (3 by 3 grid of voxels) are calculated where the central voxel is the target voxel. The transition probabilities for the nine-cell neighbourhood are used in the cluster optimisation step of the algorithm and will be discussed more fully in the following Section.

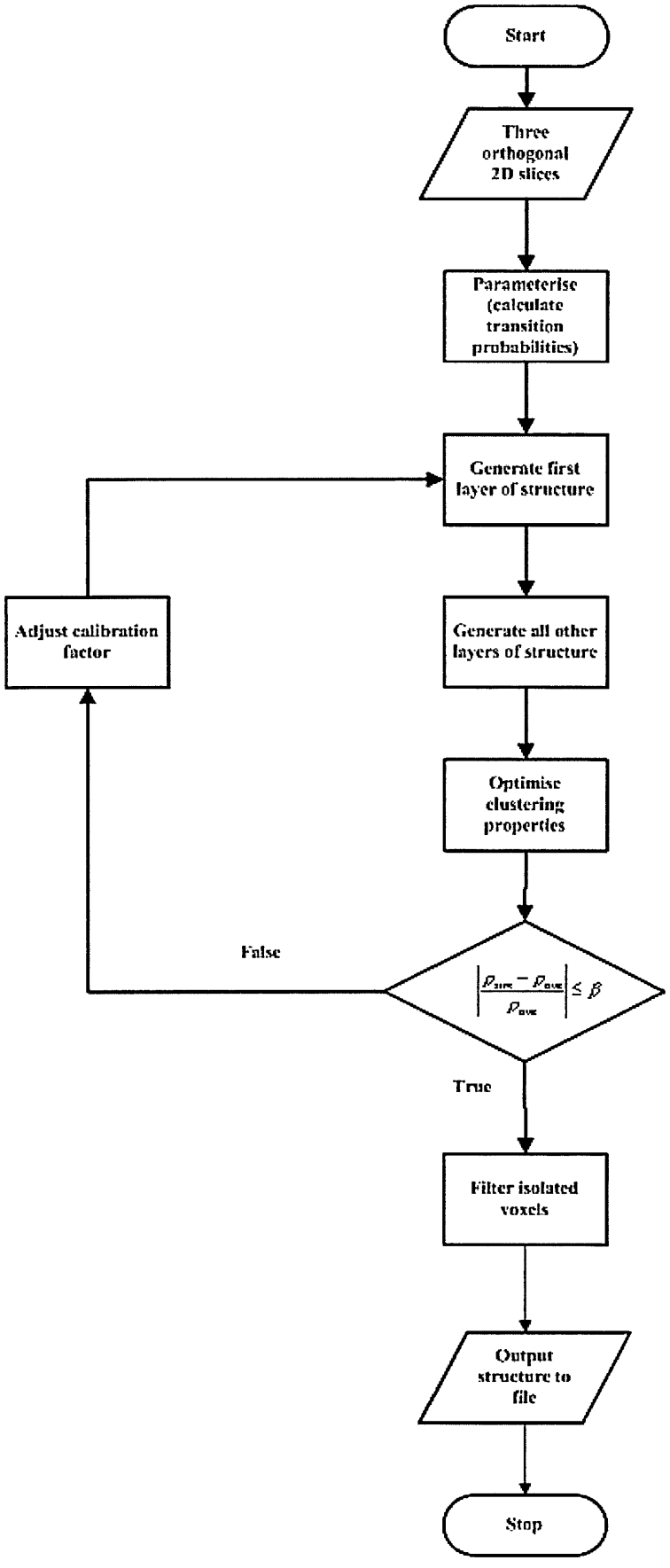


Figure 6. 1 - Flowchart of the pore architecture model algorithm

6.1.2 Generating the 3D Matrix

To initiate a structure simulation, the bottom layer is derived using the 2D algorithm of Wu et. al. (2005). This is initiated by allocating the first voxel of the first row of the first layer to be a solid state, this then allows the two-cell neighbourhood conditional transition probabilities of the horizontal image to be used to generate the remaining voxel states of first row and the first voxel of the second row. With these voxel states generated, the six-cell neighbourhood (Figure 6.2) and the conditional transition probabilities derived from the horizontal image are used to generate the states of the remaining voxels in the first layer. This is achieved by placing the six-cell neighbourhood over the known voxel states (i.e. the first voxel in the second row and the first 3 voxels in the first row) and generating the states of voxels at position (i, j) and (i, j+1) simultaneously. The state of the voxel at position (i, j) is generated based on the conditional transition probability for the combination of voxel states at positions (i-1, j-1), (i-1, j), (i-1, j+1) and (i, j-1); then the state of the voxel at position (i, j+1) is generated based on the conditional probabilities of the now known five voxel states. The six-cell neighbourhood is then progressed along the row by two voxels and the process is repeated until all voxel states in the row have been generated. A new row is initiated by generating the state of the first voxel using the two-cell conditional transition probabilities in conjunction with the state of last voxel generated on the previous row. A raster-scanning technique is implemented on each layer where the states of voxels on even rows are generated from left to right and those on odd rows are generated from right to left. This technique results in an efficient one-pass scan of each layer of the structure.

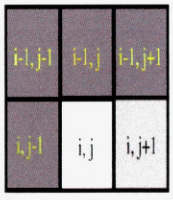


Figure 6. 2 - Six-cell neighbourhood applied to 2D slices to calculate transition probabilities and to generate first layer of structure simulation (cells i, j and i, j+1 are target voxels)

To extend the simulated structure to the next layer, the two-cell neighbourhood conditional transition probabilities derived from the vertical training image is used to allocate the state of the first cell in the new layer. The remainder of the first row of the new layer is generated in the same manner as above. However, the six-cell neighbourhood is used in a vertical direction. That is, the voxel state combination of the first three voxel in the first row of the previous layer, and the first voxel of the new layer, is considered to generate the second and third voxel states in the first row of the new layer in conjunction with the transition probabilities derived from the vertical training image. With a new layer initiated in this manner, the remainder of each subsequent layer of the 3D lattice is obtained by combining the six-cell neighbourhoods from each orthogonal section resulting in the localised fifteen-cell neighbourhood shown in Figure 6.3. The fifteen-cell 3D anisotropic neighbourhood shown in Figure 6.3 is achieved by combining the six-cell neighbourhoods in each orthogonal direction (1 vertical, 1 horizontal and 2 perpendicular). Using this shape of neighbourhood is an attempt to capture some of the structural heterogeneity found in real soils into the simulated structures. The most obvious 3D neighbourhood, a cubic neighbourhood, would not contain this critical feature.

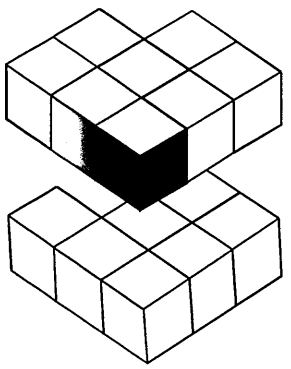


Figure 6. 3 - 15-cell neighbourhood used in generating the 3D lattice of voxels, target voxels shown in black.

The state of the two target voxels (shown in black in Fig. 6.3) are generated based on the voxel state combinations contained within the fifteen-cell neighbourhood. Applying the six-cell neighbourhood in each orthogonal direction and using the transition probabilities from the appropriate set generates a 'guess state' for each target voxel. Because the training images used to simulate a structure are not spatially referenced, there is no known formal method of combing the statistical properties from the orthogonal planes. Therefore, a pragmatic voting scheme has been developed to generate each target voxel state. The derivation of a target voxel state (T) is achieved by obtaining a guess state from each transition probability set and is combined with a further calibration state (c). Using integer arithmetic, the target voxel state is derived such that:

$$T = (h + v + p + c) / 3 \qquad (6.1)$$

Where h, v and p are the guess states from the horizontal, vertical and perpendicular neighbourhoods, respectively. Calculating T using Eqn 6.1 and integer arithmetic always produces a value of 0 or 1, representing a pore or solid state respectively.

Calibration of the model is required to increase the accuracy of the resultant porosity of the 3D structure. This is achieved by adjusting the state of the target voxel within the range of the known inputs, in this case, the average porosity of the training images. The calibration state (c) is generated as:

$$c = \alpha - \rho_{ave} + r \qquad (6.2)$$

Where α is the constant porosity calibration factor, ρ_{ave} is the average porosity of the training images and r is a random number generated for each target voxel where $0 \leq r < 1$. By using only the integer part of the result of Eqn 6.2 this calibration method always returns a value of 0 or 1, representing a vote for pore (0) or solid (1) and thus provides a means of biasing the target voxel state. Each pair of target voxels (black cells in Figure 6.3) of all subsequent layers of the 3D structure are generated in this manner using the raster scanning scheme detailed above.

The next step in the algorithm is to optimise the clustering properties of the simulated structure this is required, as the structure as it stands at this point in the simulation does not exhibit large correlated regions (either pore or solid) that are found in the training data. This is achieved by applying the nine-cell neighbourhood transition probabilities. By enlarging the neighbourhood greater clustering of both solid and pore space can be achieved. The structure is scanned in each orthogonal direction using a 2D nine-cell neighbourhood where the target voxel is in the centre of a 3*3 grid. The target voxel is updated according to the states of its nearest eight neighbours.

6.1.3 Optimising the 3D Matrix

To optimise the simulated structure an iterative process is employed which recalculates the porosity calibration factor (α) that in turn biases the voting scheme that generates the target voxel states. The porosity of the simulated structure is calculated. If this lies within a predefined range from the average porosities of the training images the modelled structure is deemed acceptable. The desired porosity range is defined by the maximum deviation parameter (β) and the average porosity (ρ_{ave}) of the training images. A simulated structure with porosity ρ_{sim}, is only accepted if:

$$\left| \frac{\rho_{sim} - \rho_{ave}}{\rho_{ave}} \right| \leq \beta \qquad (6.3)$$

If the structure does not have a porosity value within the desired range a new porosity calibration factor is derived as:

$$\alpha = \alpha + 0.1 * (\rho_{sim} - \rho_{ave}) \qquad (6.4)$$

A new 3D structure is then generated as described in the Section 6.1.2. This process is repeated until a structure within the desired porosity range is achieved.

Having produced a simulated structure with a porosity value within the desired range the final step of the model involves removing isolated solid or pore voxels. The smoothing parameters

τ_{min} and τ_{max} are the minimum and maximum clustering coefficients and are used to identify isolated voxels. Each voxel in the 3D grid is scrutinised and the states of the nearest 26 neighbouring voxels are examined. The total number of solid voxels in the neighbourhood is found. If the voxel being scrutinised is pore and the total number of solid neighbours is greater than or equal to τ_{max}, then all voxels in the 27-cell neighbourhood are updated to be solid. If the voxel under scrutiny is solid and the total number of solid neighbours is less than or equal to τ_{min}, then all voxels in the neighbourhood are updated to be pore. This final step is applied to optimise the 3D clustering properties of the simulated structure.

6.2 Model Validation

The initial model validation consisted of testing the model at the extremes. That is, could the model produce a completely porous structure if trained on all pore data and an all solid structure if using completely solid training data? Having executed these tests and obtained the desired correct results the model can be tested against real data. To validate the ability of the model to capture naturally occurring heterogeneity, structures were simulated based on 2D training image slices (1 horizontal, 1 vertical and 1 perpendicular) extracted from the CT imaged structures introduced in Chapter 5. The simulated structures are quantified using the metrics described in Chapter 5 and the results compared to those of the CT structures. These results are presented in the following section.

The sensitivity of the model to the input parameters was investigated to identify if these parameters had an effect on the efficiency and accuracy of the model. The porosity calibration factor is updated iteratively to find the optimum value for a particular model realisation, thus the starting value of this parameter could affect the length of time a model takes to complete and it is desirable that the model run-time is kept to a minimum. Sensitivity analysis was carried out on the initial porosity calibration factor (α) and a value of 0.7 was found to be the optimum starting value for these datasets. It was found that this value resulted in the minimum number of recursions within the model to achieve a structure within the desired porosity range and thus the model run-time is kept to a minimum. However, this may not

necessarily be the optimum value for the initial porosity calibration factor if different types of soil structure were to be simulated using this model.

The value for the maximum derivation parameter (β) is set to 0.03 to ensure that the difference in porosity between the simulated structures and the average porosity of the training images is not greater that 3%; this is deemed a suitable porosity deviation range for the simulated structures. However, for structures with porosity values less than 3%, the maximum derivation parameter was reduced to 0.01 to guarantee the generation of a simulated structure with porosity within 1% of the average porosity of the training images. The smoothing parameters τ_{min} and τ_{max} are constants within any model realisation. Therefore these will not impact on the model run-time but may have an effect on the accuracy of the final simulated structure. The smoothing parameters τ_{min} and τ_{max} are set to 3 and 24, respectively. These values are chosen to allow only clusters of up to 3 voxels to be filtered out. Thus keeping smoothing of the structure to a minimum and avoiding any artefacts being introduced into the simulated structure.

6.3 Results

This section details the results obtained from the quantification of the simulated structures and the comparison of these metric values with those obtained from the CT structures presented in the previous Chapter. The results presented here correspond to metric values derived using the vertices detection neighbourhood in the metric computation algorithms.

The results presented in the following sections are based on CT structures of size 100^3, and the related simulated structures of the same size. However, the same analysis has been carried out on structures of size 200^3 to test for scaling effects. The general conclusions are consistent with those presented below. The results obtained from the 200^3 structures provide further evidence that this pore architecture model is valid.

6.3.1 Porosity (ρ)

The porosity data presented in Table 6.1 shows that the modelled structures capture the variation of porosity found across the CT-imaged sets. It can be seen that the simulated structures are not significantly different ($P > 0.05$) from the CT-imaged samples in terms of porosity for all sample sets. However, although not significant there is a general trend to underestimate the porosity of the simulated structure compared to the CT-imaged data sets.

Due to the Pore Architecture Model being calibrated on the average porosity of the training images, the resultant porosity of the simulated structures are expected to be good. However, the training images must be representative of the structure being modelled in order to yield accurate simulations (see Section 6.1.1).

Metric	Porosity		Surface Connected Porosity		Vugh Porosity		% Pore Space Surface Connected	
	CT	Sim	CT	Sim	CT	Sim	CT	Sim
Sample 1 - Zero-day	5.1 (0.6)	4.4 (0.9)	2.9 (1.0)	1.4 (0.5)	2.2 (0.3)	3.0 (0.5)	50.1 (12.4)	26.1 (6.0)
Sample 2 - Control	5.7 (1.0)	5.4 (1.7)	4.0 (1.0)	3.09 (2.0)	1.7 (0.2)	2.3 (0.5)	67.4 (5.7)	39.9 (18.5)
Sample 3 - Bulk soil	8.1 (2.4)	7.9 (1.7)	5.8 (2.0)	5.2 (2.1)	2.3 (0.4)	2.7 (0.5)	64.9 (9.1)	53.0 (14.2)
Sample 4 - Rhizosphere	15.1 (5.4)	14.6 (3.7)	13.1 (4.0)	12.9 (4.1)	2.1 (0.6)	1.7 (0.4)	74.1 (12.2)	76.7 (10.8)

Table 6. 1 - Comparison of porosity metrics for the simulated structures against the CT-imaged structures using the Vertices detection neighbourhood. Numbers in brackets are the standard error values.

6.3.2 Surface Connected Porosity (ρ_{sc})

The surface connected porosity values (Table 6.1) obtained from the simulated structures were not significantly different ($P > 0.05$) from the CT-imaged structures for all sample sets. Again, although not significant, there is a general trend to underestimate the surface connected porosity in the simulated structures.

6.3.3 Vugh Porosity (ρ_v)

The vugh porosity is defined as the fraction of the total structure that is pore and is unconnected to the outer environment. The average vugh porosity values for both the CT and simulated structures are shown in Table 6.1. It can be seen that there is no significant difference ($P > 0.05$) in the metric values obtained across the entire sample sets. Furthermore, it can be seen in Figure 6.4 that when the vugh porosity (ρ_v) is plotted against the air filled porosity (ρ), the CT and simulated structures shown similar trends across the porosity range. The increasing scatter at low porosities may be partially attributed to the fact that lower porosity structures are probably less representative of the real bulk soil of a larger soil sample, therefore the correlation between the CT and modelled points may be affected at lower porosities.

6.3.4 Fraction Of Pore Space Surface Connected (γ)

These metric values can be seen in Table 6.1. It is shown that the fraction of pore space surface connected for the low porosity structures are considerably lower than the values of the corresponding CT structures. However, a paired t-test of the metric values showed no significant differences ($P > 0.05$) between the CT and simulated structures across all data sets.

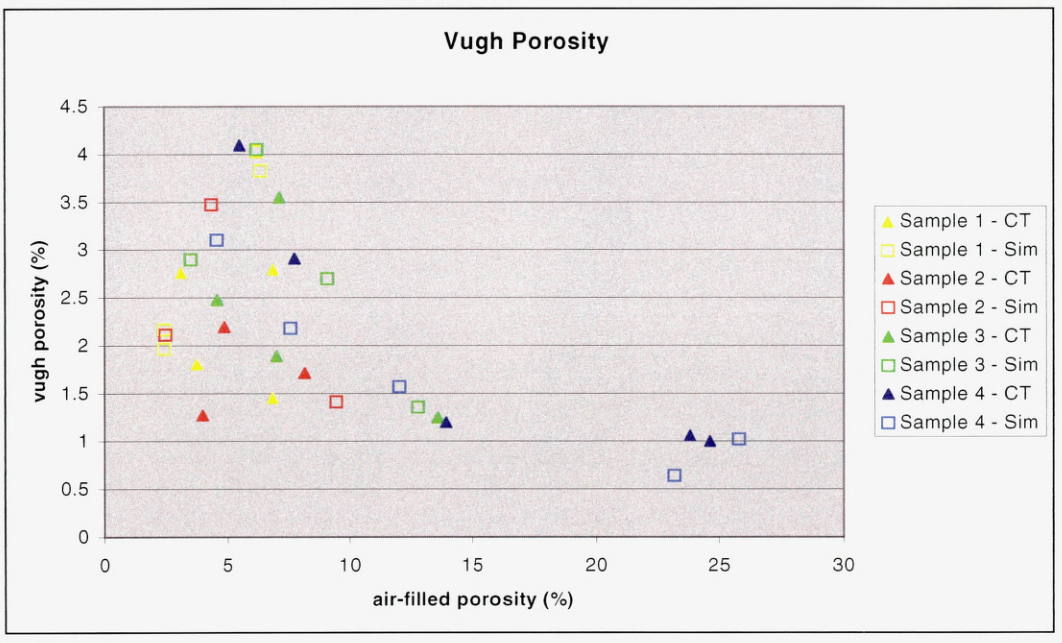


Figure 6. 4 - Vugh porosity of CT-imaged and simulated structures.

6.3.5 Solid Voxel Distance Distributions

For each CT-imaged sample the frequency distribution of the distance to nearest connected pore for each voxel was calculated. It was found that higher porosity structures (sample sets 3 and 4) resulted in higher maximum distances to nearest connected pore than found in lower porosity structures (sample sets 1 and 2). It can be seen that there are a greater number of voxels with distances of less than 50 microns to a connected pore in the high and medium porosity samples than in the low porosity samples (Figure 6.5). This is a consequence of the higher surface connected porosities of these structures.

The frequency distributions obtained from the simulated structures follow similar qualitative trends to the CT-imaged structures. With the sample sets 3 and 4 (high porosity) showing approximately half the maximum distance to nearest connected pore than sample sets 1 and 2 (low porosity). Thus, where the surface connected porosity values are higher, the simulations yield higher numbers of voxels with distances of less than 50 microns to a connected pore than structures simulated with lower porosities.

The agreement between the distributions for the simulated and CT structures is better for short distances than it is for larger distances. There is no general trend towards over or underestimation as porosity increases. Larger distance measures will be sensitive to the simulated distribution of smaller pores and there is evidence from Figure 6.5 that the model is weakest in the prediction of the small-scale pore space.

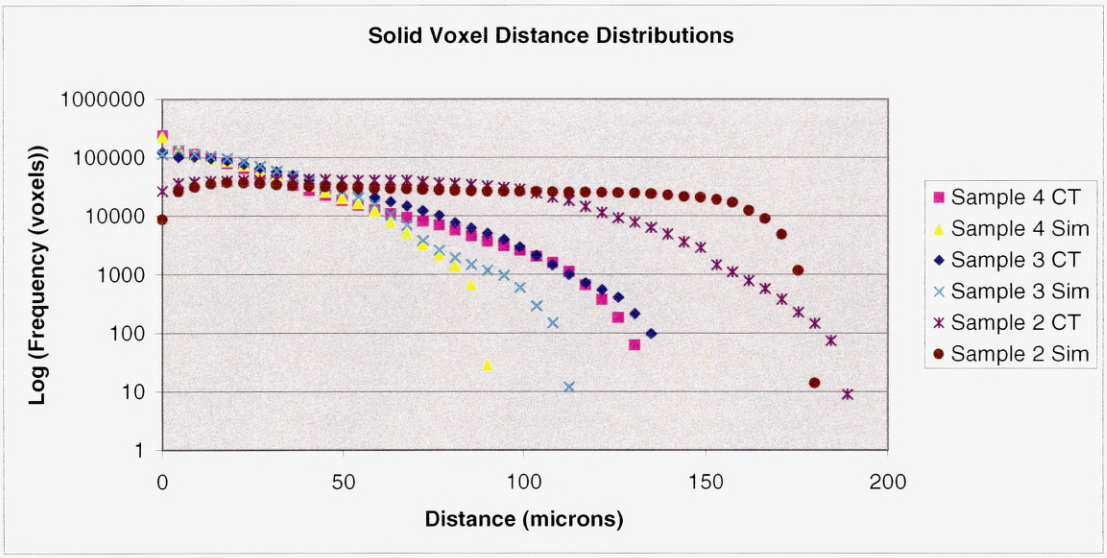


Figure 6. 5 - Solid voxel frequency distance distributions for low (sample 2), medium (sample 3) and high porosity (sample 4) structures showing both CT and simulated distributions.

6.3.6 Fractal Dimension (d_f)

The fractal dimensions obtained for the simulated and CT-imaged structures are shown in Table 6.2. It is shown that the values for the simulated structures are not significantly different ($P > 0.05$) from the CT-imaged data sets. This metric shows that the simulated structures capture the heterogeneity found in the CT-imaged samples.

Metric	Total Pore Clusters		Pore Clusters Surface Connected		% of Clusters Surface Connected	
	CT	Sim	CT	Sim	CT	Sim
Sample 1 - Zero-day	735 (50.3)	924 (65.7)	22 (3.1)	116 (4.7)	3.1 (0.004)	12.7 (0.005)
Sample 2 - Control	745 (50.9)	557 (140.6)	20 (2.3)	67 (15.9)	2.8 (0.003)	12.6 (0.01)
Sample 3 - Bulk soil	668 (37.3)	561 (107.3)	24 (3.6)	64 (10.9)	3.6 (0.003)	11.9 (0.009)
Sample 4 - Rhizosphere	787 (170.6)	401 (127.6)	33 (7.9)	55 (15.4)	4.1 (0.003)	15.2 (0.009)

Table 6. 2 - Pore complexity measurements, comparison between CT and simulated structures.

6.3.7 Spectral Dimension (d_s)

The spectral dimensions calculated for the simulated and CT-imaged structures are shown in Table 6.2. It can be seen that there is a no significant difference ($P > 0.05$) between the values obtained for the simulated and CT sample sets. However, although not significant, there is a general trend towards an under-estimation of the spectral dimensions for the simulated structures. This would indicate that the simulated structures have a less continuous and more tortuous pore space than the original CT-imaged samples. It is evident from the porosity measures presented above that the simulated structures have a less connected pore space than the CT samples and thus the lower spectral dimension values would be expected.

6.3.8 Interior Design Metrics

6.3.8.1 Total Pore Clusters

The total number of pore clusters was calculated for each of the simulated structures, the average values obtained across the samples sets are shown in Table 6.2 alongside the CT sample set values. It can be seen that there is a significant difference ($P < 0.05$) between the values of the simulated and CT sets for sample set 1 and sample set 4. However, samples set 2 and 3 showed no significant difference ($P > 0.05$) in the total pore clusters of the CT and simulated structures.

Sample set 1 comprises of low porosity structures (< 8%) and is the only sample set to significantly over estimate the total number of pore clusters. It can be seen that sample sets 2, 3 and 4 all shown an under-estimation of the total number of pore clusters and all these samples sets contain at least one individual structure that has a porosity greater than 8%. However, simulated samples of set 4, yield a significantly lower difference in this metric from the CT-imaged samples, which could be attributed to the fact this sample set contains greater variation of porosity samples than the other data sets.

6.3.8.2 Fraction of Pore Clusters Surface Connected

The results obtained for the fraction of pore clusters surface connected are also shown in Table 6.2. These metric values also show a significant difference ($P < 0.05$) between the simulated and CT-imaged sample sets. All simulated structures show approximately 400% increase in surface connected pore clusters compared with the CT samples. This means that the simulated structures contain higher numbers of small surface connected clusters compared to fewer but larger connected porous regions in the CT structures. This metric provides further evidence that the model does not capture the small-scale connectedness of the pore space found in the CT samples.

6.3.8.3 Surface Area / Volume Distribution

The surface area / volume distribution of the pore clusters is the final metric used in the quantification of soil structure. This metric is used as a means of investigating the size and shape of the individual pore clusters. For each individual pore cluster the log of the volume against the log of the surface area is plotted as shown in Figure 6.6. The distributions obtained from the simulated structures show reasonable correspondence with the CT-imaged samples. However, it can be seen that at low porosity values the simulated structures do not capture the large pore cluster (i.e. high volume and high surface area) that is evident in the corresponding CT structure. Furthermore, the high and medium porosity simulations show more than one high volume pore cluster that is seen in the CT samples. Again this highlights the failure of the model to capture the connectedness of the pore structure via small pathways.

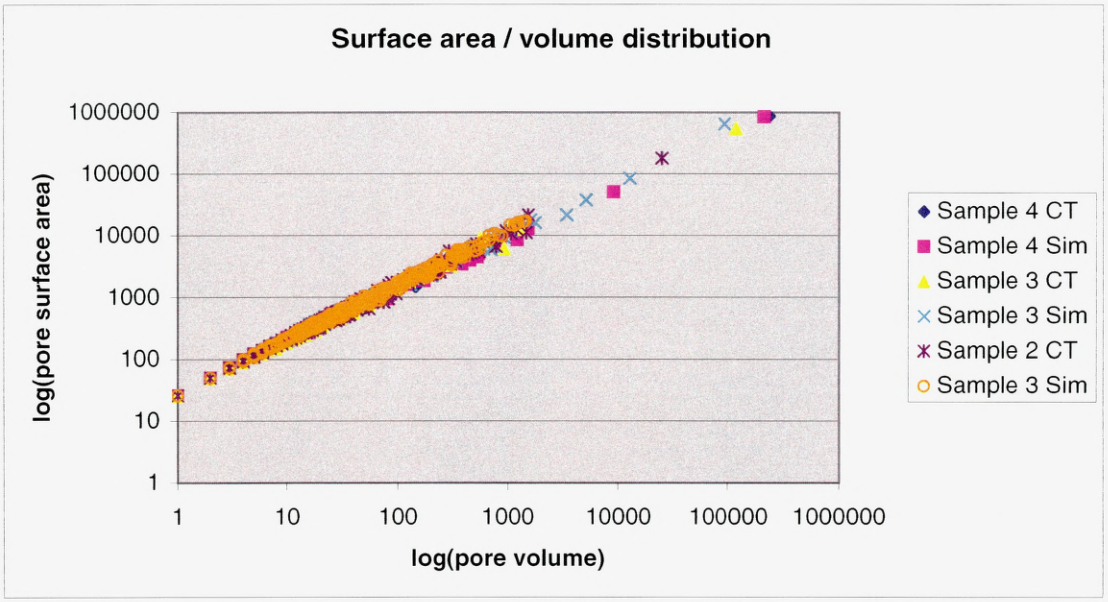


Figure 6. 6 - Surface area / volume distribution for a high (sample 4), medium (sample 3) and low (sample 2) porosity structures, simulated and CT structure distributions shown.

6.4 Discussion

The results presented above show that the model produces simulated structures that adequately capture the variation in structure found across the CT-imaged structures. On average, the simulations performed well across all the sample sets. However, the analysis showed that when comparing individual simulated structures against the original CT-imaged structures there was a distinct split between high porosity (> 8%) and low porosity (< 8%) samples. The simulations performed well for the high porosity structures. For the samples with a porosity of less than 8%, the simulated structures replicated the bulk porosity. However, these low porosity simulated structures were poor at replicating the surface connected porosity, the vugh porosity and the fraction of pore space surface connected. This is particularly evident from the visual representations shown in Figure 6.7, which are the simulated structures corresponding to the CT-imaged structures shown in Figure 5.7 in the previous Chapter. It can be seen that the high and medium porosity structures show good correspondence with CT-imaged structures both in the metric values and visually. However, the low porosity simulated structure, although capturing the bulk porosity as shown in Figure

6.7 (a), the pore space is far less correlated than the CT-imaged structure. This is most obvious in Figure 6.7 (b), where the surface connected pore space is shown. It can be seen that the surface connected porosity value is significantly lower than the CT-imaged value and the visual representation also shows the surface connected pore space is fragmented in comparison to the large correlated pore structure seen in the original CT-imaged structure (Fig. 5.7 (b)).

It is shown in Horgan and Ball (2005) that soil structures modelled using a Boolean random set model display similar trends to those found in the structures presented here. Horgan and Ball (2005) show that the proportion of total pore space that is unconnected to the outer environment increases linearly as the porosity increases until about 8% porosity after which the vugh porosity rapidly decreases. It is shown that this peak of vugh porosity at around 8% porosity can be related to the percolation threshold. The percolation threshold indicates the porosity at which the pore space appears continuously connected throughout the structure. An intensive simulation study by Lorenz and Ziff (2001) has shown that the percolation threshold occurs at about 6.5% porosity. It can be seen that the samples studied here produce results that are in agreement with these previous investigations; with the CT-imaged samples showing the percolation threshold to be at 5.49% porosity and the simulated structures threshold is seen at a porosity of 6.2%.

The soil voxel distance distributions for the simulated samples show quantitative agreement across high porosity samples however at low porosity (< 8 %) show only qualitative agreement with the CT-imaged sets. As this distribution relates to the distribution of oxygen in the soil matrix, the correspondence between the distributions from the simulated and CT structures (see Fig 5.7 (d) and 6.7(d)), particularly for smaller distances, suggests that the model should be able to capture some aspects of the functionally important characteristics of high porosity structures. However, at larger distances the model tends to be poor. The most likely reason for these discrepancies is the inability of the model to simulate rare events, particularly the appearance of large-scale correlated pore structure in low porosity media.

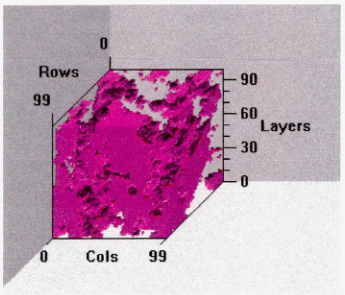
The metrics that measure pore complexity give further support to the validity of the pore architecture model presented in this Chapter. The metric values obtained for the fractal dimension of the simulated structures show no significant difference from the values obtained for the CT-imaged structures. This indicates that the scaling properties of the pore volume found in the CT structures are captured within the simulated structures. The spectral dimension of the simulated structures shows the model has a general trend to underestimate the continuity of the connected pore space. However, as there is no significant difference found between spectral dimensions calculated for the CT and simulated structures this shows that the pore architecture model sufficiently captures the tortuosity of the pore networks. Therefore, the simulated structures should posses similar transport properties, in terms of gas and solute transport through the pore space, as the CT-imaged structures.

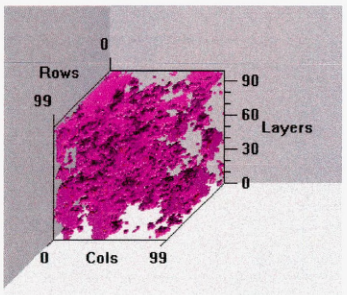
The quantification of the pore space via the interior design metrics shows that in general, the model produces structures that capture the overall surface connected clusters more accurately at high porosities than low porosities. However, it is apparent there are significantly more individual pore clusters connected to the outer environment in the simulated structures that the CT-imaged samples. This highlights the fact that the model does not capture the small connecting pore channels that exist in the CT-imaged samples across all porosities. This is particularly evident in the comparison of the low porosity simulated structure found in Figure 6.7 (c) with the corresponding CT structure shown in Figure 5.7 (c). It can be seen that the simulated structure lacks the relatively large connected pore cluster found in the CT-imaged sample. This provides further evidence of the inability of the model to capture the rare relatively isolated pores that are found in the CT-imaged structures especially in low porosity samples. This can be attributed to the fact that the model is dependent on the conditional probabilities derived from the training images, and if the small connecting pore channels are not captured the resultant simulated structure will lack these features.

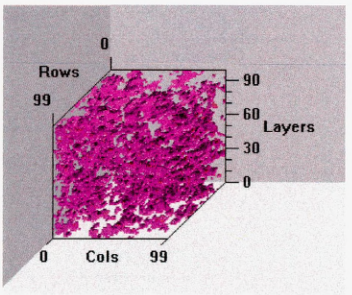
Overall, there is a general trend across all metrics where higher porosity simulated structures show better correlation with the CT-imaged structures than lower porosity sets. The work presented here has used CT-imaged soil structures to validate a method of simulating the 3D

soil habitat. It has been shown that the model adequately captures the pore architecture of structures where the porosity is above 8%. Below this value the simulated structures capture the main porosity values but not the detailed characteristics of porosity such as connectivity and clustering properties. The reason for this failure is due to the inability of the model to simulate the rare events such as large pores in low porosity structures. This can be attributed to the fact that low porosity structures provide less data relating to the connectedness of the pore space.

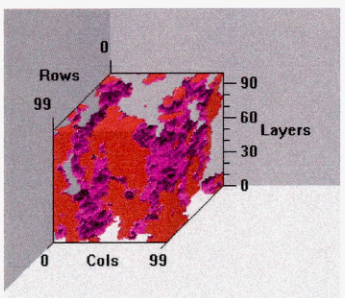
(a) (i) ρ	23.15%	12.81%	4.34%
(ii) d_f	2.83	2.96	2.95

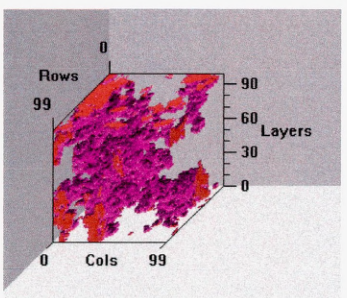


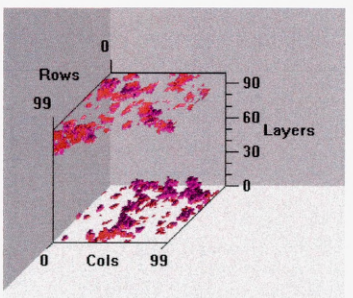




(b) (i) ρ_{sc}	22.51%	11.45%	0.87%
(ii) γ	97.22%	89.39%	20.09%
(iii) d_s	1.59	1.43	1.29

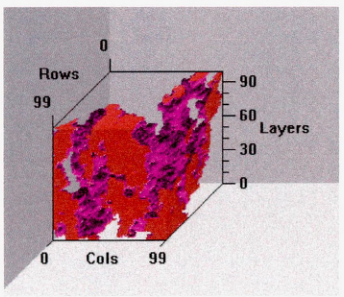


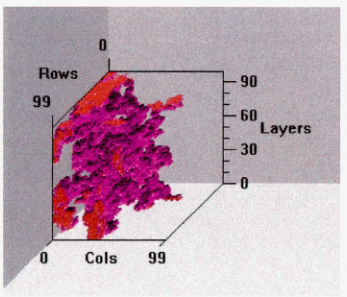


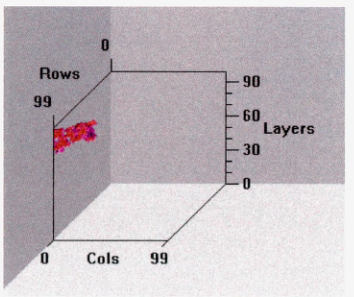


(c) Single surface connected pore cluster

(i) Cluster Volume	21.2%	9.31%	0.12%







(d) Solid voxel distance distributions

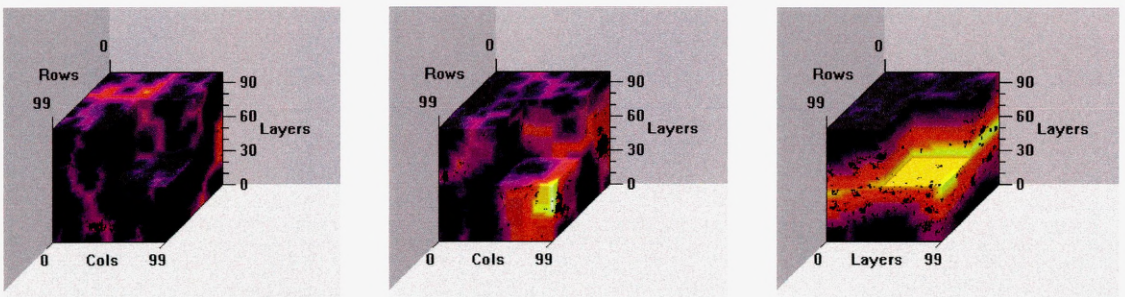


Figure 6. 7 - Visual representation of pore scale properties for a high, medium and low porosity simulated structures.

Chapter 7 – Modelling Soil Functional Characteristics

7.1 Introduction

One of the aims of this thesis is to investigate the impact of the physical structure on the functional characteristics of soil. As discussed in Chapter 2, the soil-microbe complex provides essential services to agriculture, waste management, the water industry and also plays an important role in global carbon dynamics. Therefore, an understanding of the key processes and the resultant impact of these processes are crucial to developing sustainable management regimes.

The microbes contained within the soil microenvironment are key facilitators in a number of biological processes that occur within the soil system (Dorioz, et. al, 1993, Feeney et. al, 2006). However, the spatial distribution of the microbes within a soil system may affect the system-scale functional characteristics. In addition, the previous Chapters have shown that the physical structure of soil can vary greatly in terms of the geometry of the pore and solid matrix as described by the metrics introduced in Chapter 5. Because the geometry of soil structure influences the transport of resources around the soil microenvironment (Rappoldt 1992) it is expected that the differing structural geometries will also result in different functional characteristics. Therefore, an understanding of whether the soil structure, or the distribution of microbes within it, determines the bioavailability of diffusible and soluble compounds in soil is sought. It is important to know under what conditions structure dominates over the microbial distribution and vice versa. To examine this relationship between structure and microbes the processes of oxygen transport and microbial respiration is modelled. The following sections detail how the modelling framework has been extended to incorporate the functional characteristics of oxygen transport and microbial respiration.

7.2 Specification of Soil Functional Characteristics

The incorporation of functional characteristics into the Soil Architecture Model (SAM) is non-trivial. The assumptions and definitions relating to the functional characteristics must be

documented and transparent in order for the model to be justifiable and hence useful in terms of studying the emergent system-scale behaviour of the soil microenvironment. The SAM is an individual based model and thus all functional characteristics are defined at the individual level of the model components thus allowing the system-scale properties to emerge as a consequence of interactions between the components parts. The specification and assumptions relating to the functional characteristics of the SAM are listed in Table 7.1; these were informed from the review of the literature as presented in Chapter 2.

1	Solid matrix of structure is assumed to be comprised of solid particles and small pores containing water (i.e. is assumed saturated)
2	The transport of oxygen is modelled using a diffusion process
3	A solid voxel may contain many microbes
4	A pore voxel may contain no microbes
5	Microbes may be randomly distributed around the solid-pore interface or randomly within the solid matrix
6	All microbes respire at the same constant rate
7	Microbes may only respire if the oxygen concentration of the containing voxel is greater than, or equal to, the number of microbes contained within the voxel, multiplied by the microbe respiration constant
8	A voxel is assumed anoxic if the oxygen concentration is less than the total number of microbes within the voxel multiplied by the microbe respiration constant
9	The supply of oxygen to the structure is provided by a layer of source oxygen at the top and bottom planes of the structure
10	All remaining boundaries are treated as reflective (i.e. zero flux boundaries)

Table 7. 1 - Specification of functional characteristics

7.2.1 Gas Diffusion

The transportation of oxygen through a structure is simulated using a diffusion process. The diffusion process is shown to mimic the random motion that gas particles exhibit (Crank, 1975). The diffusion equation in three-dimensions is:

$$\frac{\partial c}{\partial t} = D_i\left(\frac{\partial^2 c}{\partial x^2}+\frac{\partial^2 c}{\partial y^2}+\frac{\partial^2 c}{\partial z^2}\right) \qquad (7.1)$$

where c is the oxygen concentration, t is time, x, y, z, are the spatial dimensions and D_i is the diffusion coefficient of the pore or solid phase. This equation is continuous in both space and time. A numerical approximation is used to model Eqn 7.1, which is the FTCS (forward time centred space) finite-difference explicit method (Crank, 1975). This method of approximation discretizes the diffusion equation and enables the spatial oxygen concentration within the physical domain (i.e. the soil structure) to be treated as a 3D grid and the oxygen concentrations at each successive time step are calculated from known values from the previous time step.

The discretization of the continuous equation (Eqn 7.1) requires the replacement of the differential operator ∂ with the discrete analogue Δ. By applying a Taylor's series over the spatial domain and holding the time domain constant we can obtain the 2nd order central difference approximation of the 2nd derivative. This is followed by a Taylor's series over the temporal domain while holding the spatial domain constant. Thus, giving the discretized approximation of Eqn 7.1 as:

$$\frac{u^{m+1}_{x,y,z}-u^{m}_{x,y,z}}{k} = \frac{D_i}{3h^2}\left[u^{m}_{x+h,y,z}+u^{m}_{x-h,y,z}+u^{m}_{x,y+h,z}+u^{m}_{x,y-h,z}+u^{m}_{x,y,z+h}+u^{m}_{x,y,z-h}-6u^{m}_{x,y,z}\right]$$

where u is the concentration, x, y and z correspond to the spatial dimensions, the superscript m represents the temporal scale, h is the grid spacing unit in all spatial dimensions and k is the time step unit. Upon rearranging, this becomes:

$$u_{x,y,z}^{m+1} = u_{x,y,z}^{m}\left[1-\frac{6kDi}{h^2}\right]+\frac{kDi}{h^2}\left[u_{x+h,y,z}^{m}+u_{x-h,y,z}^{m}+u_{x,y+h,z}^{m}+u_{x,y-h,z}^{m}+u_{x,y,z+h}^{m}+u_{x,y,z-h}^{m}\right]$$

The discretization process results in the stability criterion:

$$\frac{kD}{h^2} < \frac{1}{6} \qquad (7.2)$$

The stability and accuracy of the diffusion process relies on Eqn 7.2 being fulfilled. If this stability criterion is not fulfilled the resultant finite difference approximation becomes unstable. This instability can be seen in the example presented by Crank (1975, Example 8.4.1) where the finite difference solution bears no resemblance to the analytical solution when the stability criterion is not met. The implementation of this diffusion process in to the SAM is validated against this known analytical solution; the details are presented in Section 7.3.1.

7.2.2 Microbial Respiration

The continuous equation that models the diffusion-respiration process is:

$$\frac{\partial c}{\partial t} = D_i\left(\frac{\partial^2 c}{\partial x^2}+\frac{\partial^2 c}{\partial y^2}+\frac{\partial^2 c}{\partial z^2}\right)-Q \qquad (7.3)$$

where Q is a zero-order sink term that describes the respiration of the system. The respiration rate is expressed in units of gas consumed per unit time per unit volume of soil. Due to the individual-based nature of the model, the respiration rate must be defined at the level of the individual microbe and is consistent with Eqn. 7.3. The discretized version of Eqn 7.3 is:

$$u_{x,y,z}^{m+1} = u_{x,y,z}^{m}\left[1-\frac{6kDi}{h^2}\right]+\frac{kDi}{h^2}\left[u_{x+h,y,z}^{m}+u_{x-h,y,z}^{m}+u_{x,y+h,z}^{m}+u_{x,y-h,z}^{m}+u_{x,y,z+h}^{m}+u_{x,y,z-h}^{m}\right]-s_{x,y,z}q$$

where s is the number of microbes at grid position x, y, z and q is the respiration constant for a single microbe.

7.2.3 Microbial Distribution

To investigate the influence of microbial distribution on the system-scale soil functionality, the distribution of microbes is specified by two distinct distributions. A fixed number of microbes are assumed to be randomly distributed either:

- On the pore-solid interface

or

- Within the solid matrix

By using a fixed biomass and distributing the individuals differently within a structure, it makes it possible to compare the results obtained from the simulations using different structures. In accordance with the specification detailed in Table 7.1, only solid voxels within a structure may contain microbe objects. Thus, with the ability of solid voxels to contain many microbe objects, a fixed number of microbes can be distributed within the desired solid voxel set (i.e. interface or matrix voxels).

7.2.4 Quantification of Functional Characteristics

To allow system-scale functionality of soil structures to be compared, the functional characteristics must be quantified. Because this project is focused on oxygen availability and microbial respiration, the metrics used to analyse the simulations are the oxic fraction of the solid matrix and bulk respiration. The oxic fraction of the solid matrix is used to assess the availability of oxygen within a structure and the bulk respiration is used to measure the level microbial life that can be sustained.

It is important to specify exactly how these metrics are calculated as assumptions that may be made about the definition of these measurements can have implications on the analysis and conclusions drawn from the resultant values. Due to the object oriented nature of the modelling framework the bulk metrics are calculated from the individual component parts.

These metrics will change over time, therefore they must be calculated at the end of every time step of the simulation so overall trends in individual structures can be analysed. The oxic fraction of the solid phase of the structure is computed as:

$$\text{Oxic solid fraction} = 1 - (\text{number of anoxic solid voxels} / \text{total solid voxels}),$$

with anoxic voxels defined as in Table 7.1. Again, because of the design of the modelling framework it is possible to calculate the oxic or anoxic fraction of any phase of the soil in a similar manner. However, it is crucial that the logic that sets the anoxic state of voxels is fully defined and documented.

The bulk respiration is calculated as the total number of microbes that respired during the time step in which the value is being computed. Average values of both these metrics may also be calculated when the steady state has been reached. This is necessary because it is likely that final steady state of the diffusion-respiration simulations will contain some oscillation over time as the oxygen moves through the structure to areas of demand.

7.3 Implementation and Validation

The development of the SAM modelling framework to include the functional characteristics as defined in Table 7.1 required the inclusion of the Diffusion and Microbe classes and some extension to the Structure class (Fig 3.2). The Diffusion class encapsulates the mathematical process of the finite difference method detailed in Section 7.2.1. The parameters used to drive the diffusion process such as diffusion coefficients and discretization parameters are defined in the Diffusion schema of the Plang specification file and accessed by the main program using the PlangSchema attribute of the Diffusion class. The respiration parameter is also defined in the Plang specification file in the Microbe schema.

The object-oriented approach coupled with the Plang system provides a formal and coherent strategy for identifying, defining and documenting model parameters. Furthermore, this design

and implementation strategy provides the modelling framework with the ability to run multiple diffusion processes concurrently, with the specific parameters for each process defined independently through the Plang specification file. In the same manner, the characteristics of different groups of microbes can be defined using Plang. This decoupling of process from parameters is potentially a powerful modelling strategy that may be fully exploited in future work.

7.3.1 Validation of the Diffusion-Respiration Process

To facilitate the systematic testing and validation of the diffusion process implementation against input data with known analytical solutions, output from the model is written to file that documents all results, including intermediate results that are not necessarily used in the final model analysis. A spreadsheet is then used to manually carry out calculations and allow comparison of the model results with the expected values.

The implementation of the diffusion process alone (i.e. no respiration) was validated using a simple numerical example as detailed in Crank (Crank, 1975, Example 8.4.2) where an analytical solution is also available. This example initially consists of two straight lines:

$$c = 2X, \qquad 0 \le X \le \frac{1}{2}, \qquad T = 0,$$

$$c = 2(1 - X), \qquad \frac{1}{2} \le X \le 1, \qquad T = 0$$

where the two faces of the sheet are maintained at zero concentration so that:

$$c = 0, \qquad X = 0 \text{ and } 1, \qquad T \ge 0$$

This problem is symmetric around x = ½, therefore the numerical solution need only be calculated for $0 \le X \le ½$. The analytical solution to this problem is:

$$c = \frac{8}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \sin \frac{1}{2} n\pi \sin n\pi X \exp(-n^2 \pi^2 T)$$

Calculating the values of c at X = 0.3 and X = 0.5 with h = 0.1 and k = 0.001, thus satisfying the stability criterion of Eqn 7.2, it is shown in Table 7.2 that the solution obtained using the

SAM implementation of the finite-difference method shows good correspondence with the analytical solution.

	X = 0.3			X = 0.5		
T	A.S	F.D.S	Percentage difference	A.S	F.D.S	Percentage difference
0.005	0.5966	0.5970	0.07	0.8404	0.8553	1.77
0.01	0.5799	0.5812	0.22	0.7743	0.7773	0.39
0.02	0.5334	0.5321	0.24	0.6809	0.6745	0.94
0.10	0.2444	0.2393	2.09	0.3021	0.3067	1.52

Table 7. 2 - Comparison of numerical and analytical solution. A.S. = analytical solutions, F.D.S = finite-difference solution, Percentage difference calculated as a percentage of the analytical solution.

The combined diffusion-respiration process implementation was validated against the steady state example presented in Rappoldt (1992, Chapter 2). Rappoldt (1992) shows that the differential equation describing the steady state diffusion-respiration process for a plane sheet is:

$$\frac{d}{dx}\left[-D\frac{dc(x)}{dx}\right]=-Q \qquad 0\le x\le 1$$

with boundary conditions:

$$c(x)=1 \text{ for } x=0 \text{ and } x=1, \quad \text{and} \quad \frac{dc(x)}{dx}=0 \text{ for } x=\frac{1}{2}$$

where c is the concentration, x is the depth of the structure and Q is the respiration rate and D is the diffusion coefficient. The analytical solution to this problem is:

$$c=1-\frac{Q}{D}\left[\frac{x}{2}-\frac{x^2}{2}\right]$$

This is in essence a homogeneous problem where there is no structure (i.e. all voxels are of the same state) and each voxel contains one microbe. The SAM was used to simulate this problem using a homogenous structure of 50^3 voxels all assumed to be solid, with the diffusion coefficient, $D = 1$ and the respiration rate of a single microbe, $Q = 10$. The graph in Figure 7.1 shows that the steady state analytical solution is reached using the numerical method as detailed in Section 7.2.2. It can be seen that the analytical solution shows negative concentrations, this is due to the analytical equation depends on the gradient of the oxygen concentration rather than the absolute value. Whereas, it can be seen that the numerical solution has been adapted to force the solution to go to zero which is a valid requirement of the modelling framework, as in reality there cannot be negative oxygen concentrations. Because of this the numerical solutions has some variation from the analytical solution .

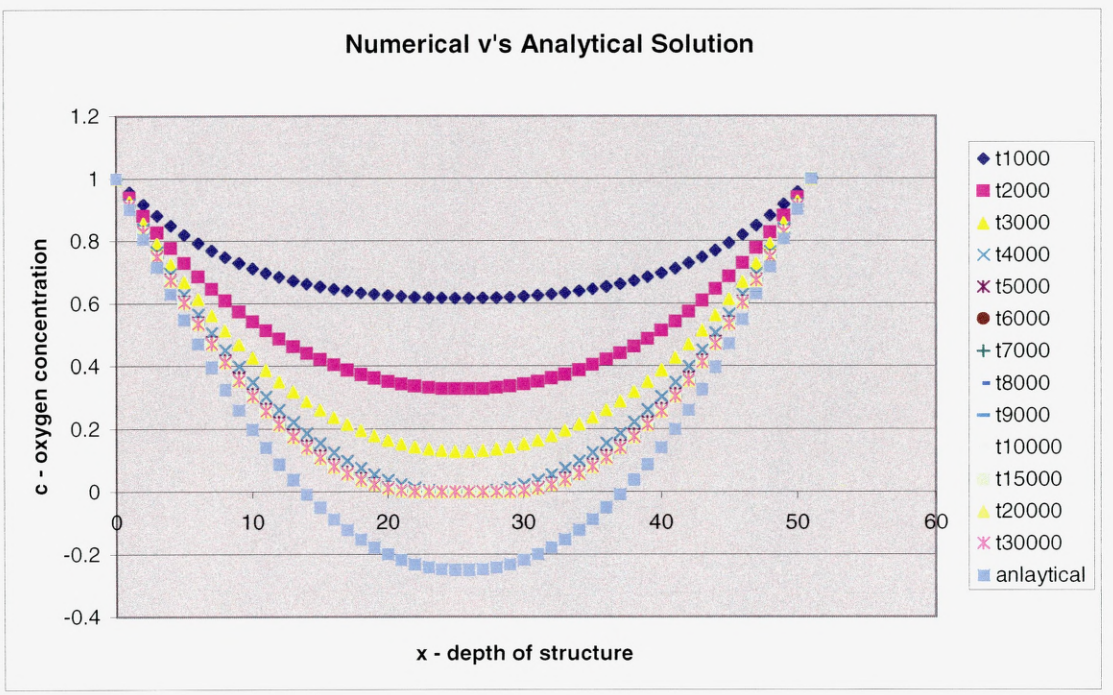


Figure 7. 1 - Comparison of numerical and analytical solution for diffusion-respiration process.

7.4 Results

This section details the outcome of simulations realised using the SAM model with the inclusion of the functional characteristics as described in the previous sections. Due to time constraints and the amount of processing time required to complete these simulations, small structures (10^3 voxels) were used. These structures were obtained by randomly sampling from a CT structure and the corresponding simulated structure. This allows comparison between the CT and simulated structures in terms of functional characteristics.

Ten random samples are taken from both the CT and corresponding simulated structure and the resultant structures spanned a range of porosities and hence different structural characteristics. Each structure is used as an individual microcosm in which to simulate oxygen transport and microbial respiration with the microbes randomly distributed either within the solid matrix or at the pore-solid interface. At the start of every simulation the structure is assumed to be full of oxygen with the diffusion coefficient of the solid voxels set to 10,000 smaller than pore voxels, a fixed biomass of 800 microbes is used with each microbe having an individual respiration rate of 10 units per second. The discretization of the diffusion-respiration parameters is carried out internally within the SAM with checks made to ensure stability of the numerical solution. The diffusion-respiration process is allowed to run until a steady state is achieved, in that the flux of oxygen from the source layers into the structure had reached equilibrium.

7.4.1 The Impact of Structure on Function

To investigate the impact of the physical soil structure on the functional characteristics, each structure is seeded with the fixed biomass distributed randomly within the solid matrix voxels. The final respiration of the samples can be found in Figure 7.2. The respiration values are averaged over the final 1000 time steps of the simulation to account for the inherent fluctuation in the number of microbes that may be respiring at any single time step. It can be seen in Figure 7.2 that the bulk respiration of the structures decreases with increasing porosity. This is expected due to fewer solid voxels available in the high porosity samples,

which results in a more densely distributed microbes. This leads to an increase in the local demand for oxygen, which in turn is consumed over shorter timescales by the microbes. It can be seen that the oxic fraction of the solid matrix decreases linearly with increasing porosity as shown in Figure 7.3, again the values shown are average values over the final 1000 time steps of the simulation.

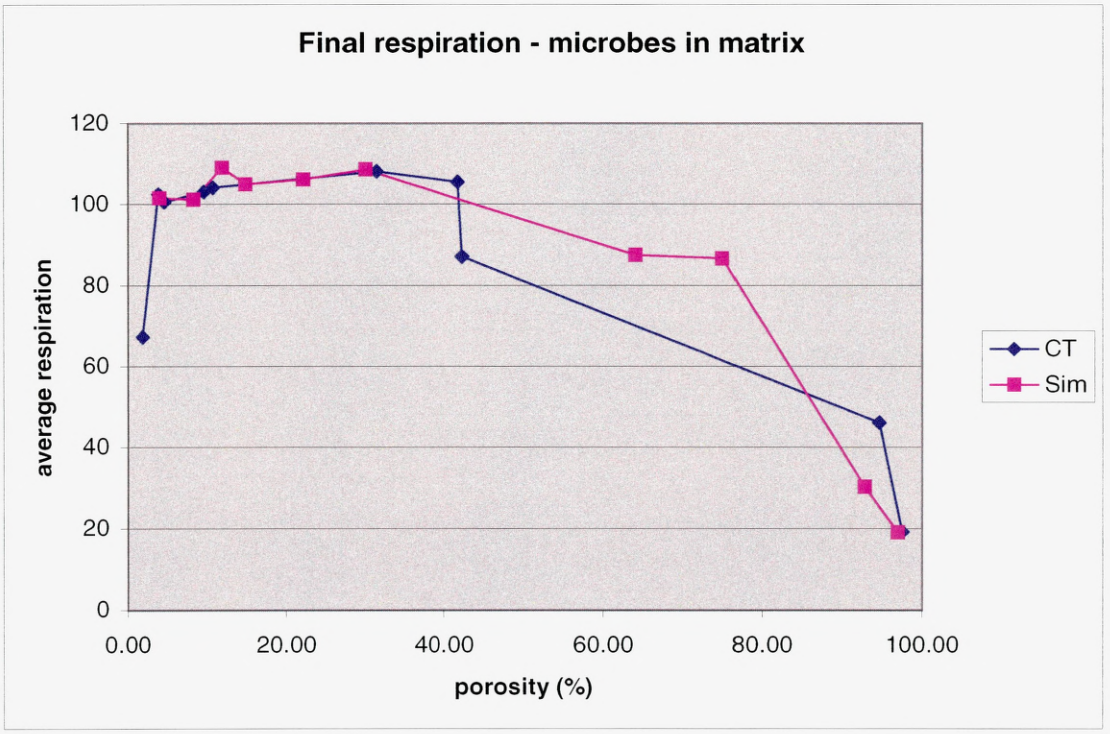


Figure 7. 2 - Final bulk respiration with microbes distributed in the solid matrix, CT and simulated structures shown.

The final bulk respiration values obtained using the simulated structures with microbes distributed within the solid matrix (Fig. 7.2) show a similar trend to the CT structures with the high porosity structures yielding the lower bulk respiration. The oxic fraction of the solid matrix is also shown to decrease linearly as porosity increases as expected (Fig. 7.3).

Both the CT and simulated structures exhibit an optimum respiration in structures with porosity values between 15% and 35%. This can be attributed to the relatively short distance that solid voxels are from oxygen filled pore voxels in these structures. Furthermore, the

microbes will be more uniformly distributed throughout the solid voxels in these samples, as there are more solid voxels available for population by microbes than in the higher porosity samples where the microbe density per voxel will be higher.

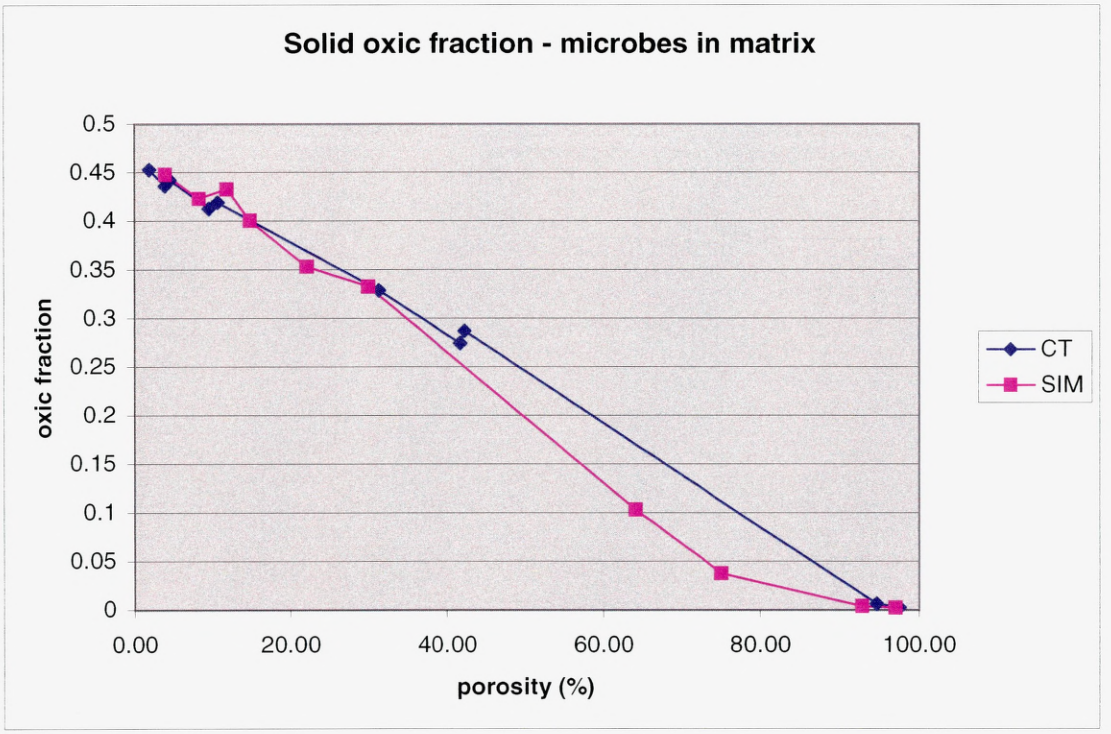


Figure 7. 3 - Oxic fraction of solid matrix with microbes distributed in solid matrix, CT and simulated structures shown.

7.4.2 The Impact of Microbial Distribution on Function

To investigate the impact of microbial distribution on the functional characteristics of a soil sample, the same structures were used as in the previous section. However, in this case the fixed biomass is randomly distributed around the pore-solid interface, with microbe objects residing in the solid interface voxel. There are inherently fewer solid voxels available for microbes to reside in with this distribution that in the pervious section. Therefore, more microbes will be allocated to individual solid interface voxels. Because there are fewer available voxels for the microbes to reside in, the populated voxels in this experiment are likely to contain a greater number of microbe objects than in the pervious experiment.

Figure 7.4 shows the average bulk respiration values obtained with the microbes distributed on the pore-solid interface for both the CT and simulated structures. It can be seen that both data sets exhibit the same trend where low and high porosity structures yield the lowest respiration with an optimum respiration found between 30% and 50 % porosity. The simulated structures show good correspondence with the CT samples, indicating that the Pore Architecture Model captures the key features of soil structure relevant to soil aeration.

The oxic fraction of the solid matrix, as shown in Figure 7.5, again yields a decrease as porosity increases. However, with the microbes distributed at the interface the oxic fraction is shown to be almost twice that found when the microbes were distributed within the solid matrix. This suggests that with microbes distributed at the interface the supply of oxygen to the microbes comes from the pores rather than being drawn from the surrounding soil. This is justified as the diffusion of oxygen through the solid matrix is much slower than in the pores.

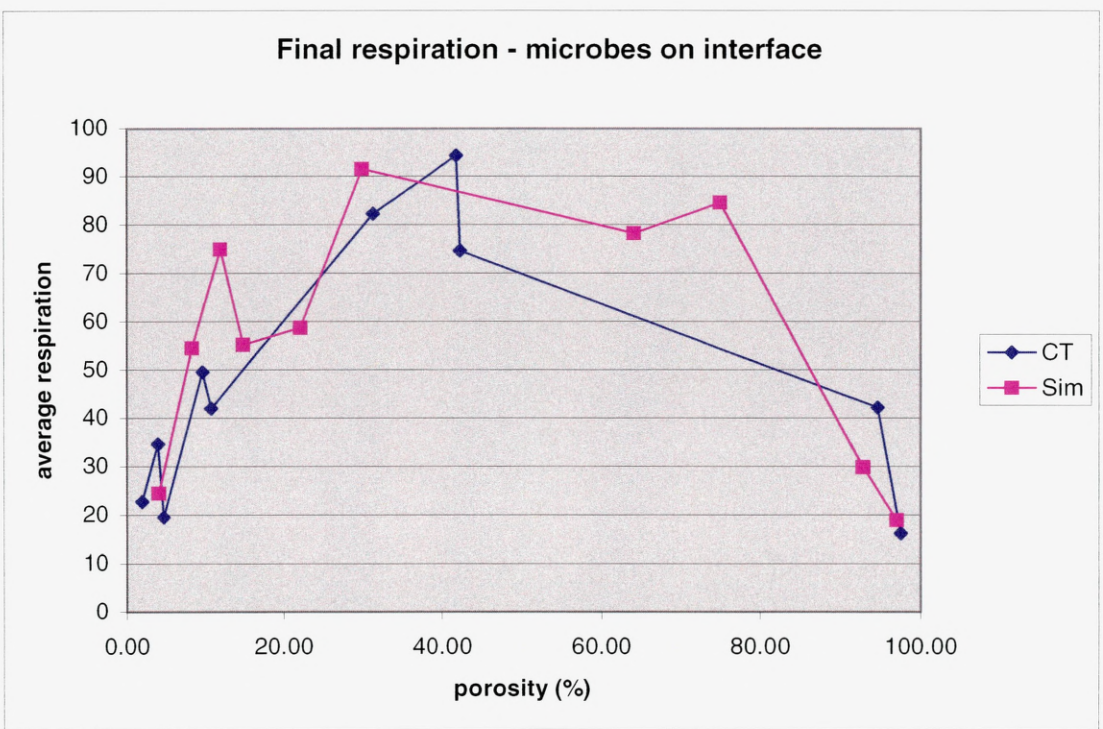


Figure 7. 4 - Final bulk respiration with microbes distributed at the pore-solid interface, CT and simulated structures shown.

The respiration trends shown in Figure 7.4 are more pronounced than those found in Figure 7.2 where the microbes were distributed throughout the solid matrix. In general, the bulk respiration is lower when the microbes are distributed on the interface as opposed to within the solid matrix. However, this may be an artefact of the implementation of the uptake of oxygen by microbes due to the fact that the modelling framework only allows respiration to take place within a voxel if there is enough oxygen for all microbes within the voxel to respire. This is a severe limitation of the modelling framework at present because there may be enough oxygen diffusing into the voxels to allow some of the microbes to respire, thus the structure may be able to sustain a higher bulk respiration than is found in these experiments.

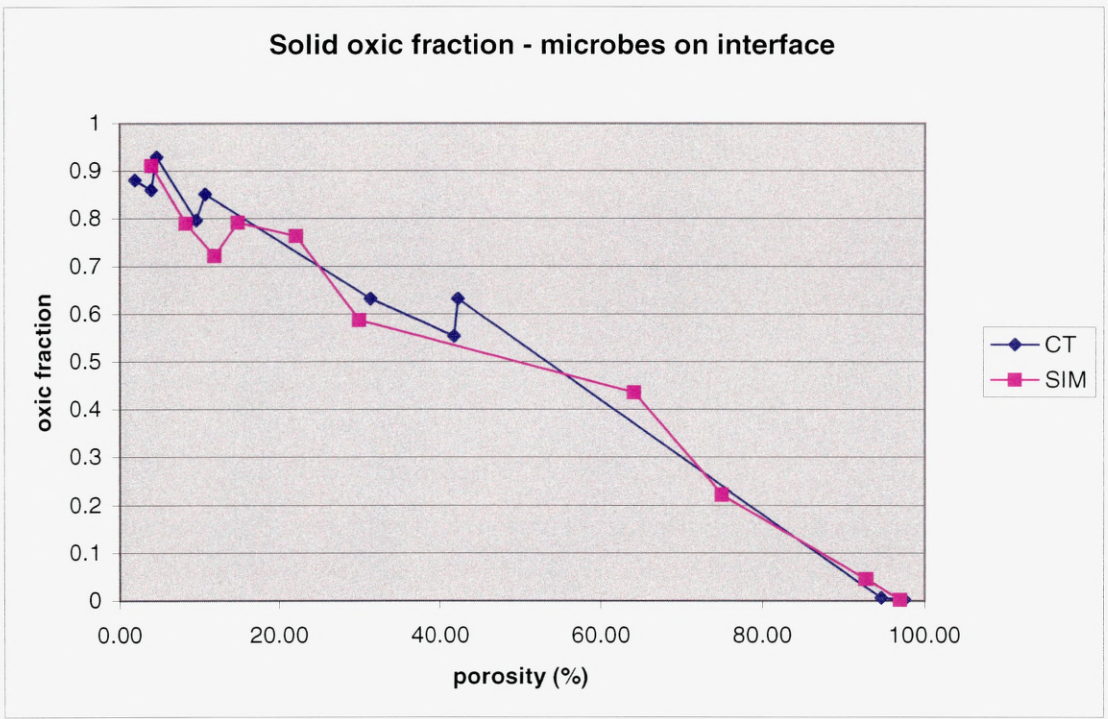


Figure 7. 5 - Oxic fraction of solid matrix with microbes distributed on pore-solid interface, CT and simulated structures shown.

7.5 Discussion

The aim of this Chapter is to investigate the impact of the physical structure of soil on the system-scale functional characteristics. To achieve this, the SAM has been extended to include an implementation of a diffusion-respiration process with facilities to allow the seeding of microbe objects within a structure and the inclusion of metrics to facilitate the analysis of these simulations. The implementation of these tools into the modelling framework is non trivial and the specification and validation of these aspects are crucial to obtaining accurate and informative results. The specifications of the functional aspects of the modelling framework were informed from the literature review presented in Chapter 2. However, the specifications were kept as simple as possible to avoid the introduction of artefacts that may not necessarily be understood and visible within the resultant simulations. A model is only as good as the assumptions on which it is based; therefore the documented and validated specifications are an intrinsic part of the model that provide confidence in the results obtained from the simulations.

The key aim of the modelling framework in terms of modelling functional characteristics is to allow system-scale properties to emerge as a consequence of the interactions of the component parts. Due to the object-oriented design of the Soil Architecture Model (SAM) this was a natural extension and relatively straight forward to implement. Because all components of the model, such as voxel and microbe objects, maintain their own state and behaviour the individual traits of the components interact on a local scale thus allowing the system-scale properties to emerge.

This Chapter shows how this individual based strategy is exploited to examine the relationship between soil structure and the distribution of microbes in relation to system-scale functional characteristics. Soil structures that span a range of porosities were seeded with microbes randomly distributed either within the sold matrix or at the pore-solid interface. It was found that in general, the simulations with microbes distributed within the solid matrix maintained higher respiration that the simulations where the microbes were distributed at the interface. Both sets of simulations showed optimum bulk respiration in medium porosity

structures. However, the simulations with microbes in the solid matrix exhibited this optimum bulk respiration in structures with porosities between 15% and 35% whereas the simulations with microbes distributed at the interface yielded optimum bulk respiration in structures with porosities between 30% and 50%. Furthermore, the results of the simulations presented in this Chapter show that the structures simulated using the Pore Architecture Model (PAM) show good correspondence with the CT structures in terms of bulk respiration and oxic fraction. This provides further evidence that the PAM captures the key features of soil structure that are important in regulating oxygen transport and microbial respiration.

The methodology presented in this Chapter provides evidence that individual based modelling of the soil-microbe complex is appropriate and applicable. However, the functional characteristics as defined in Table 7.1 are limited in terms of the investigative power of the modelling framework. Firstly, the definition of microbial respiration within the modelling framework is a simplistic view of reality, which in essence creates an all or nothing situation. However, this definition of respiration provides a base line from which the mechanism of microbial respiration can be extended and enhance to produce a more realistic model. In addition, the seeding of microbes using the two distributions defined in Section 7.2.3 is also limited but again provides a base line on which to build. The ability to incorporate real microbe distributions extracted and calculated from soil samples would be the ideal solution. However, to date, these data are only available in 2 dimensions thus a method to extrapolate to 3 dimensions must be developed. Furthermore, because the diffusion-respiration process implementation within the SAM is computationally intensive it was not possible to simulate the diffusion-respiration process on the larger structures presented in the preceding two Chapters. Therefore, it is not possible to link the full set of structural quantification metrics detailed in Chapter 5 to the functional characteristics defined in this Chapter. With the use of distributed computer networks it would be possible to overcome this problem.

In conclusion, the Soil Architecture Model (SAM) provides for the first time a spatially explicit, individual based theoretical framework on which to investigate the soil-microbe system in terms of the relationships between structure and function. The SAM provides a software

architecture that can be reused and further extended to provide a virtual laboratory for investigation and experimentation of the soil-microbe system.

Chapter 8 – Summary and Future Work

8.1 Summary

The main aim of this thesis has been to investigate the influence of physical soil structure on the functional characteristics. To enable this investigation a software architecture has been developed that encapsulates the theoretical modelling framework described in the preceding Chapters. A review of the relevant literature explored the different aspects of the soil microenvironment with particular emphasis on soil structure and the influence that physical structure has on biological activity. The review revealed the inherent complexity of the soil-microbe system and highlighted the importance of including both the biotic and abiotic elements of the microenvironment into a model as the spatial heterogeneity of these elements combined affects the system-scale functional characteristics. Existing models of the soil microenvironment tend to model either the abiotic phase such as soil structure coupled with transport characteristics (Blunt 2001, Wu, 2006), or the biotic phase such as bacterial colony growth (Ginovart et. al, 2002, Kreft, 1998). The novel aspect of the modelling framework developed during this doctoral programme is that both the physical structure and biotic elements are combined into a single architecture. For the first time, the Soil Architecture Model (SAM) provides a spatially explicit individual based 3D model of micro-organisms in soil.

The SAM provides a virtual laboratory that facilitates the investigation of the 3D soil microenvironment. The SAM provides a suite of metrics to quantify the physical 3D soil structure, as well as an implementation of a model that simulates 3D structures from 2D data. Furthermore, the SAM provides a means of simulating oxygen diffusion, microbial distributions and microbial respiration in three dimensions. The development of the SAM into a software architecture using an object-oriented design strategy supports both flexibility and extensibility of the modelling framework. The object-oriented strategy is a natural methodology for designing and implementing individual-based models as the state and behaviour of individual model components are encapsulated within classes. Each class can

be viewed as a model component type that can be used to create multiple copies of that type within a model realisation. This design and implementation strategy yields a software architecture where the data (i.e. state) of individual model components are protected within objects that in turn control the process of a model realisation. This is an important concept in individual based modelling, as the behaviours and interactions that model components exhibit are driven by the state and behaviour of components in the local vicinity. However, this allows system-scale properties and behaviours to emerge as a consequence of the individual components and this is key to studying the complex soil-microbe system.

All individual based models are inherently parameter driven thus the Plang system (Milne, 2006) is incorporated into the modelling framework. Plang provides a formal method to identify, define and document model parameters. But more importantly, the use of Plang decouples model parameters from the model process in that all parameter values are defined outside the modelling framework in the Plang specification file with access to the values facilitated through the Plang API. The Plang specification requires parameters to be defined in schemas that can be directly related to the classes within the main model thus aiding the identification of parameters required for any model component. This combination of object-oriented design and Plang is potentially a powerful design methodology for developing individual based models in the future.

The development of the modelling framework using this methodology facilitated with ease the implementation of the structure quantification metrics. The quantification metrics implemented within the SAM incorporate a range of quantification metrics from basic porosity measures to more complex methods such as fractal dimensions and the interior design metrics (Section 5.3). The literature review revealed the many and wide-ranging methods currently used to characterise soil structure. Methods range from lab-based quantification of pore-size distributions through to morphological and fractal characterisation. A number of software packages have been developed that allow the characterisation of soil in terms of hydraulic properties, such as DXSoil (Delerue and Perrier, 2002) and 3DMA-Rock (Lindquist (n.d)). However, no such tool exists that supports the quantification of soil structure in terms of

aeration properties. The suite of metrics available through the SAM provides a coherent package of quantification tools that allow a range of soil structure properties to be characterised. The quantification of the physical soil structure is necessary to enable the comparison of the micro-structure of different soils and furthermore facilitate the relationship between structure and function to be investigated. However, the functional characteristics that are under study dictates the type of quantification metrics that are practical. For the purposes of this thesis, the metrics capture properties of the structure that are important in gaseous transport.

The Pore Architecture Model (PAM) presented in Chapter 6 provides a model to allow the simulation of 3D soil structures from 2D data. The simulated structures obtained using the PAM were parameterised using 2D data extracted from the known 3D CT structures introduced in Chapter 5 thus allowing the simulated structures to be compared to the original structures via the quantification metrics. Wu et. al (2006) showed that this model reconstructed the 3D characteristics of a range of relatively homogeneous rock and stone and one soil sample. The investigation carried out in Chapter 6 shows that our model implementation captures the variation in structure found across the CT imaged samples. For structures with porosities above 8% the simulations show no significant differences in metric values. However, the model fails to reproduce the complex structure of soil when the porosity is below 8%. Although the bulk porosity, on which the model is calibrated, is realised in these low porosity structures, large-scale correlated pore networks are not captured.

Due to time constraints and the amount of processing power required, the simulations of soil function were limited to small randomly samples structures taken from the CT and simulated structures. This meant a full analysis of the PAM in terms of replicating functional aspects of soil structure could not be fully validated. However, it was shown that the simulated samples did replicate trends found across the CT samples in terms of bulk respiration and oxic fraction thus providing some evidence that the modelled structures produced by the PAM could be used in further investigations of soil functional characteristics.

During the functional simulation investigation the limitations of the SAM framework became apparent. Firstly, to simulate a diffusion process to steady state in a structure any larger than about 20^3 voxels requires massive processing times. This is due to the numerical method used to simulate the diffusion process within the SAM architecture. The explicit finite difference method requires the stability criterion (Eqn. 7.2) to be fulfilled and this severely limits the discretization parameters and thus results in long convergence times. A number of other numerical methods that do not have such severe restrictions are available and the implementation of these into the modelling framework could allow a more flexibility to be introduced into the functional simulation activities of the SAM.

Furthermore, the efficiency of the implementation of the software architecture was not a consideration in the design of the model. For example data types have not been optimised to minimise storage space or processing time. Parts of the implementation use integers where bytes could have been used. This may have implications for scalability of the model in terms of representing very large soil structures. However, the computation algorithms used in calculating the quantification metrics have been design and implemented with processing efficiency as a priority as discussed in Section 5.4.2.

In conclusion, the software architecture that has been developed can be reused, refined and extended to incorporate new features as well as exploit the current functionality. However, the efficiency of the implementation must be considered as a priority for future extensions of the model. Nevertheless, the investigations presented in this thesis show that the underlying modelling strategy is sound and provides evidence that the strategy shows promise for the design, development and maintenance of individual based models of the soil-microbe complex.

8.2 Future Work

The modelling framework presented in this thesis is encapsulated within a software architecture that is reusable but more importantly extensible. Therefore, a number of

extensions to the framework are proposed which would allow a more complete model of the micro-scale functional characteristics of the soil-microbe complex. The main areas of extension to the modelling framework are:

- The modelling framework requires the implementation of the Crank-Nicholson method of diffusion (Crank 1975) to more timely diffusion processes to be simulated. The inclusion of this numerical method would also enable multiple diffusion processes to be simulated concurrently within any single model realisation. This is an important attribute of a modelling framework of soil function as gases and solutes require different diffusion parameters and thus separate calculation processes.

- The inclusion of a water phase within the visible pore space is also a valid extension to the modelling framework. This would allow the simulation of solute transport in saturated and unsaturated soils and thus the impact of soil moisture content with respect to functional characteristics can be examined.

- An important and non-trivial extension to the modelling framework is to include mechanisms to model the structural genesis that is evident within the micro-structure of soil. Feeney et. al. (2006) and Cosentino et. al. (2006) show that soil structure at the micro-scale is modified and altered by physical and chemical processes mediated by microbial activity. Therefore, extending the functional traits of the microbes within the modelling framework to include exudates processes, coupled with mechanisms to allow the physical structure to evolve as a consequence of the collective microbial behaviour would provide the basis with which to examine the relationship between microbial exudation processes and physical structure genesis.

- Finally, the software architecture requires evaluation and optimisation in terms of data types and computation algorithms to allow extendibility of the modelling framework to larger problems i.e. the scalability of the framework requires investigation.

References

Anderson, A. N., McBratney, A. B. and Crawford, J. W. 1998. Applications of fractals to soil studies. In: Sparks, D. L., eds. *Advances in Agronomy*, 63: pp1 – 76. Academic Press, San Diego.

Antoch, J., Deshouillérs, J. and Purnaba, G. P. 1998. Revisiting the pseudorandom number generator ran1 from the numerical recipes. *Computational Statistics & Data Analysis*. 27: pp.487-495

Baveye, P., Parlange, J. and Stewart, B. A. 1998. Fractals in soil science. *Advances in Soil Science.* CRC Press, Boca Raton

Blunt, M. J. 2001. Flow in porous media – pore-network models and multiphase flow. *Current Opinion in Colliod & Interface Science*, 6: pp 197 – 207

Carr, J. R. 2003. Simple random number generation. *Computers and Geosciences*. Article in Press

Chenu, C. 1993. Clay- or sand-polysaccharide associations as models for the interface between micro-organisms and soil: water related properties and microstructure. *Geoderma*, 56: pp 143-156

Cooke, J. 1998. The beos random number libabry. [Online]. Available from: http://www.bebox.dircon.co.uk/rng.htm. [Accesses 15th November 2005]

Cosentino, D. Chenu, C. and Le Bissonnais, Y. 2006. Aggregate stability and microbial community dynamics under drying-wetting cycles in a silt loam soil. *Soil Biology and Biochemistry,* 38: pp 2053 – 2062

Couture, R. and L'Ecuyer, P. 1995. Linear recurrences with carry a uniform random number generators. 1995 Winter Simulation Conference, Hyatt Regency Crystal City, Arlington, VA, pp.263-267

Couture, R. and L'Ecuyer, P. 1997. Distribution properties of multiple-with-carry random number generators. *Math. Comput..* 66 pp.591-607.

Coveyou R. R. and Macpherson R. D. 1967. Fourier Analysis of Uniform Random Number Generators. Journal of the ACM (JACM), v.14, p.100-119

Crank, J. 1975. The mathematics of diffusion, 2nd Edition. *Clarendon Press*, Oxford

Crawford, J. W., Sleeman, B.D. and Young, I. M. 1993a. On the relation between number-size distributions and the fractal dimension of aggregates. *Journal of Soil Science,* 44: pp 555 – 565

Crawford, J. W., Ritz, K. and Young, I. M. 1993b. Quantification of fungal morphology, gaseous transport and microbial dynamics in soil: and integrated framework utilising fractal geometry. *Geoderma,* 56: pp 157-172

Crawford, J. W. 1994. The relation between structure and the hydraulic conductivity of soil. *European Journal of Soil Science*, 45: pp 493 – 502

Crawford, J. W., Matsui, N. and Young, I. M. 1995. The relation between the moisture-release curve and the structure of soil. *European Journal of Soil Science*. 46: pp 369-375

Crawford, J.W., Harris, J.A., Ritz, K. and I.M. Young. 2005. Towards an evolutionary ecology of life in soil. *TRENDS in Ecology and Evolution*, 20: pp:81-87.

Dathe, A. and Thullner, M. 2005. The relationship between fractal properties of solid matix and pore space in porous media. *Geoderma*, 129: pp 279 – 290

Delerue, J. and Perrier, E. 2002. DXSoil, a library for 3D image analysis in soil science. *Computers & Geoscience*, 28: pp 1041-1050

Dorioz, J. M. Robert, M., and Chenu, C. 1993,.The role of roots, fungi and bacteria on clay particle oragnisation. An experimental approach. *Geoderma*, 56: pp: 179 – 194

Dwyer Jr., G. P. and Williams, K. B. 2003. Portable random number generators. *Journal of Economic Dynamics & Control.* 27 pp.645-650

Entacher, K., Schell, T. and Uhl, A. 2001. Optimization of random number generators: Efficient search for high quality lcgs. *Probabilistic Engineering Mechanics.* 16 pp.289-293

Feeney, D. S., Crawford, J. W., Daniell, T., Hallet, P. D., Nunan, N., Ritz, K, Rivers, M. and Young, I. M. 2006. Three-dimensional mircoorganization of the soil-root-microbe system. *Microbial Ecology.* 52: pp 151-158

Ginovart, M., Lopez, D. and Valls, J. 2002. INDISIM, an individual-based deiscrete simulation model to study bacterial cultures. *Journal of Theoretical Biology*, 214: pp 305 – 319

Ginovart, M., Lopez, D. and Gras, A. 2005. Individual-based modelling of microbial activity to study mineralization of C and N and nitrification process in soil. *Nonlinear Analysis: Real World Applications,* 6: pp 773 – 795

Goresky, M. and Klapper, A. 2003. Efficient multiple-with-carry random number generators with maximal period. *ACM Transactions on Modelling and Computer Simulation.* 13(4): pp.310-321

Grimm, V. and Railsback, S. F. 2005. Individual-based modelling and ecology. *Princeton University Press*, Oxfordshire

Hellekalek, P. 1998. Good random number generators are (not so) easy to find. *Mathematics and Computers in Simulation.* 45 pp.485-505

Horagn, G. W. and Ball, B.C. 1994. Simulating diffusion in a Boolean model of soil pores. *European Journal of Soil Science*, 45: pp483 – 491

Horgan, G. W. and Ball, B. C. 2005. Modelling the effect of water distribution and hysteresis on air-filled pore space. *European Journal of Soil Science.* 56: pp 647-654

Janke, W. 2002. Pseudo random numbers: Generation and quality checks. *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms, Lecture Notes.* 10 pp.447-458

Johnson, A., Roy, I. M., Matthews, G. P. and Patel, D. 2003. An improved simulation of void structure. *European Journal of Soil Science*, 54: pp 477 – 490

Kao, C. and Tang, H. C. 1997. Upper bounds in spectral test for multiple recursive random number generators with missing terms. *Computers Math. Applic.* 33 pp.119-125.

Kao, C. and Tang, H. C. 1998. Several extensively tested multiple recursive random number generators. *Computers Math. Applic.* 36 pp.129-136.

Kao, C. and Wong, J. Y. 1998. Random number generators with long period and sound statistical properties. *Computers Math. Applic.* 36 pp.113-121.

Khalil, K., Mary, B. and Renault, P. 2004. Nitrous oxide production by nitrification and denitrification in soil aggregagtes as affected by O_2 concentration. *Soil Biology & Biochemistry,* 36: pp 687 – 699

Knuth, D. E. 1973. Sorting and searching. *The art of computer programming.* Vol 3. Addison-Wesley, Reading, Massachusetts

Knuth, D. E. 1981. *The art of computer programming: Seminumerical algorithims.* Second Edition ed. Reading Massachusetts: Addison-Wesley

Kreft, J., Booth, G., and Wimpenny, W. T. 1998. BacSim, a simulator for individual-based modelling of bacterial colony growth. *Microbiology,* 144: pp 3275 – 3287

Kreft, J., Picioreanu, C., Wimpenny, W. T. and van Loosdrecht, M. C. M. 2001. Individual-based modelling of biofilms. *Microbiology,* 147: pp 2897 – 2912

Lal, R. 2003. Soil erosion and the global carbon budget. *Environmental International,* 29: pp 437 – 450

L'Ecuyer, P. 1998. Uniform random number generators. In: Medeiros, D. J., Watson, E. F., Carsom, J. S., Manivannan, M.S. ed. *Proceedings of the 1998 Winter Simulation Conference, Washington, DC, pp.97-104.*

L'Ecuyer, P. 2000a. Fast combined multiple recursive generators with multipliers of the form a = +/-2^q +/-2^r. In: Joines, J. A., Barton R. R., Kang, K., Fishwick, P.A. ed. *Proceedings of the 2000 Winter Simulation Conference, Wyndhamn Palace Resort and Spa, Orlando FL, pp.683-689.*

L'Ecuyer, P. 2000b. A new class of linear feedback shift register generators. In: Joines, J. A., Barton R. R., Kang, K., Fishwick, P.A. ed. *Proceedings of the 2000 Winter Simulation Conference, Wyndhamn Palace Resort and Spa, Orlando FL, pp.690-696.*

L'Ecuyer, P. and Andres, T. H. 1997. A random number generator based on the combination of four lcg's. *Mathematics and Computers in Simulation.* 44 pp.99-107.

L'Ecuyer and Simard 2006. TestUo1: A software library in ANSI C for empirical testing of random number generators. *Laboratoire de simulationet d'optimisation. Universitie de Montreal IRO.* Dated 26 June 2006. (http://www.iro.umontreal.ca/~simardr/)

Liang, Y. and Whitlock, P. A. 2001. A new empirical test for parallel pseudo-random number generators. *Mathematics and Computers in Simulation.* 55 pp.149-158

Levin, S. A. 1998. Ecosystems and the biosphere as complex adaptive systems, *Ecosystem,* 1: pp 431 – 436

Lindquist, W. B. n.d. *3DMA-Rock a software package for automated analysis of rock pore structure in 3-D computed microtomography images* [online]. Available from: http://www.ams.sunysb.edu/~lindquis/3dma/3dma_rock/3dma_rock.html, [Accessed 9th November 2006]

Lorenz, C. D. and Ziff, R. M. 2001. Excess number of percolation clusters on the surface of a sphere, *Physica A*, 296: pp 1 – 8

Mandelbrot, B. B. 1983. The fractal geometry of nature, *Freeman*, New York

Marsaglia, G. 1997. Diehard, a battery of tests for random number generators. [Online]. Available from the World Wide Web: URL: http://stat.fsu.edu/~geo/diehard.html.

Marsaglia, G., Zaman, A. 1991. A new class of random number generators. *Ann. Appl Prob. I,* 3,462-480

Marshal, T. J. and Holmes, J. W. 1979. Soil Physics, *Cambridge University Press*, Cambridge

Matsumoto M. and Kurita Y. 1992. Twisted GFSR generators, *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, v.2, p.179-194

Matthews. G. P., Canonville, C. F. and Moss, A. K. 2006. Use of a void network model to correlate porosity, mercury porosimetry, thin section, absolute permeability, and NMR relaxation time data for sandstone rocks. *Physical Review E,* 73 031307 (9 pages)

Milne A. C. 2006. The Plang System [online]. Available from: (http://a510690.studentweb.abertay.ac.uk/Plang/index.html), [Accessed 10th October 2006]

Minar, N., Burkhart, R., Langton, C. and Askenazi, M. 1996. The Swarm simulation system: a toolikit for building multi-agent simulations, *SFI Working Paper 96-06-042 (*http://www.santafe.edu/sfi/publications*)*

Nunan, N., Ritz, K., Rivers, M., Feeney, D. S. and Young, I. M. 2006. Investigating microbial micro-habitat structure using X-ray computed tomography. *Goederma*, 133: pp398 – 407

Pachepsky, Y. A., Crawford, J. W. and Rawls, W. J. 2000. Fractals in soil science. *Developments in soil science 27,* Elsevier Science B. V., Amsterdam

Park, S. K. and Miller, K. W. 1988. Random number generators: Good ones are hard to find. *Communications of the Acm.* 31(10): pp.1192-1201

Peat et. al. 2000. Simulation of water retention and hydraulic conductivity in soil using a 3D network, *European Journal of Soil Science*, 51: pp 65 – 79

Press, W. H. et al. 1996. *Numerical recipes in c: The art of scientific computing.* Cambridge: Cambridge University Press.

Prohkova, A. 2000. How to improve a random number generator. *Computer Physics Communications.* 124 pp.12-131

Rappoldt, C. 1992. Diffusion in aggregated soil, Doctoral thesis. Wageningen Agricultural University, Wageningen, The Netherlands

Rappoldt, C. and Crawford, J. W. 1999. The distribution of anoxic volume in a fractal model of soil. *Geoderma*, 88: pp 329 – 347

Rappoldt, C. and Crawford, J. W. 2001. An approximation for the anoxic fraction of fractal soil. 4th Eastern Canada Soil Structure Workshop, Greenhouse and Processing Crops Research Centre, Harrow, Ontario, Canada.

Schmitz, O. J. and Booth, G. 1997. Modelling food web complexity: the consequence of individual-based spatially explicit behavioural ecology on trophic interactions. *Evolutionary Ecology*, 11: pp 379-398

Simms, P. H., Yanful, E. K. 2004. A discussion of the application of mercury intrusion porosimetry for the investigation of soils, including an evaluation of its use to estimate volume change in compacted clayey soils. *Geotechnique*, 54: pp 421 – 426

Serra, J. 1982. Image analysis and mathematical morphology., *Academic Press,* London

Smith, K. A., and Mullins, C. E. 2000. Soil and environmental analysis: theory, experiment and applications. 2nd Eddition, *Marcel Dekker Incorporated*

Sobol, I. M. and Levitan, Y. L. 1999. A pseudo-random number generator for personal computers. *Computers and Mathematics with Applications.* 37 pp.33-40

Tans, P. P., Fung, I. Y., and Takahashi, T. 1990. Observational constraints on the global atmosphere CO_2 budget, *Science.* 247: pp 1431 – 1438

Tisdall, J. M., and Oades, J. M. 1982. Organic matter and water-stable aggregates in soils. *Journal of Soil Science*, 33: pp141 – 163

Tezuka, S. and L'Ecuyer, P. 1992. Analysis of add-with-carry and subract-with-borrow generators. In: Swain, J. J., Goldsman, D., Crain, R.C., Wilson, J.R. ed. Proceedings of the 1992 Winter Simulation Conference, Crystal Gateway Marriott Hotel, Arlington, VA

Vogel, H. J. 1997. Morphological determination of pore connectivity as a function of pore size using serial sections. *European Journal of Soil Science*, 48: pp 365 – 377

Vogel, H. J. and Kretzschmar, A., 1996. Topological characterisation of pore space in soil – sample preparation and digital image-processing. *Geoderma,* 73: pp 23 – 38

Vogel, H. J. 2000. A numerical experiment on pore size, pore connectivity, water retention, permeability, and solute transport using network models. *European Journal of Soil Science,* 51: pp99 – 105

Vogel, H. J. and Cousin, I. 2002. Quantification of pore structure and gas diffusion as a function of scale. *European Journal of Soil Science* 53: pp 465-473.

Vogel, H. J. and Roth, K. 2001. Quantitative morphology and network representation of soil pore structure. *Advances in Water Resources,* 24: pp 233 – 242

Wichmann & Hill. 2006. Generating Good psuedo-random numbers. *Computational Statistics & Data Anlaysis.* In Press, accepted May 2006

Wegenkittl, S. 2001. Gambling tests for pseudorandom number generators. *Mathematics and Computers in Simulation.* 55 pp.281-288

Wolf, T. and Holvoet, T. 2004. Emergence Versus Self-Organisation: Different concepts but Promising when combined. In *Engineering Self-Organising Systems: Methodologies and Applications.* pp 1-15 Springer Berlin. Eds. Brueckner. S, Giovanna, D., Karageorgos, A and R Hagpal.

Wu, K., Nunan, N., Ritz, K., Young, I. M, and Crawford, J. W. 2004. An efficient Markov chain model for the simulation of heterogeneous soil structure. *Soil Sci. Soc. Am. J.* 68: 346-351

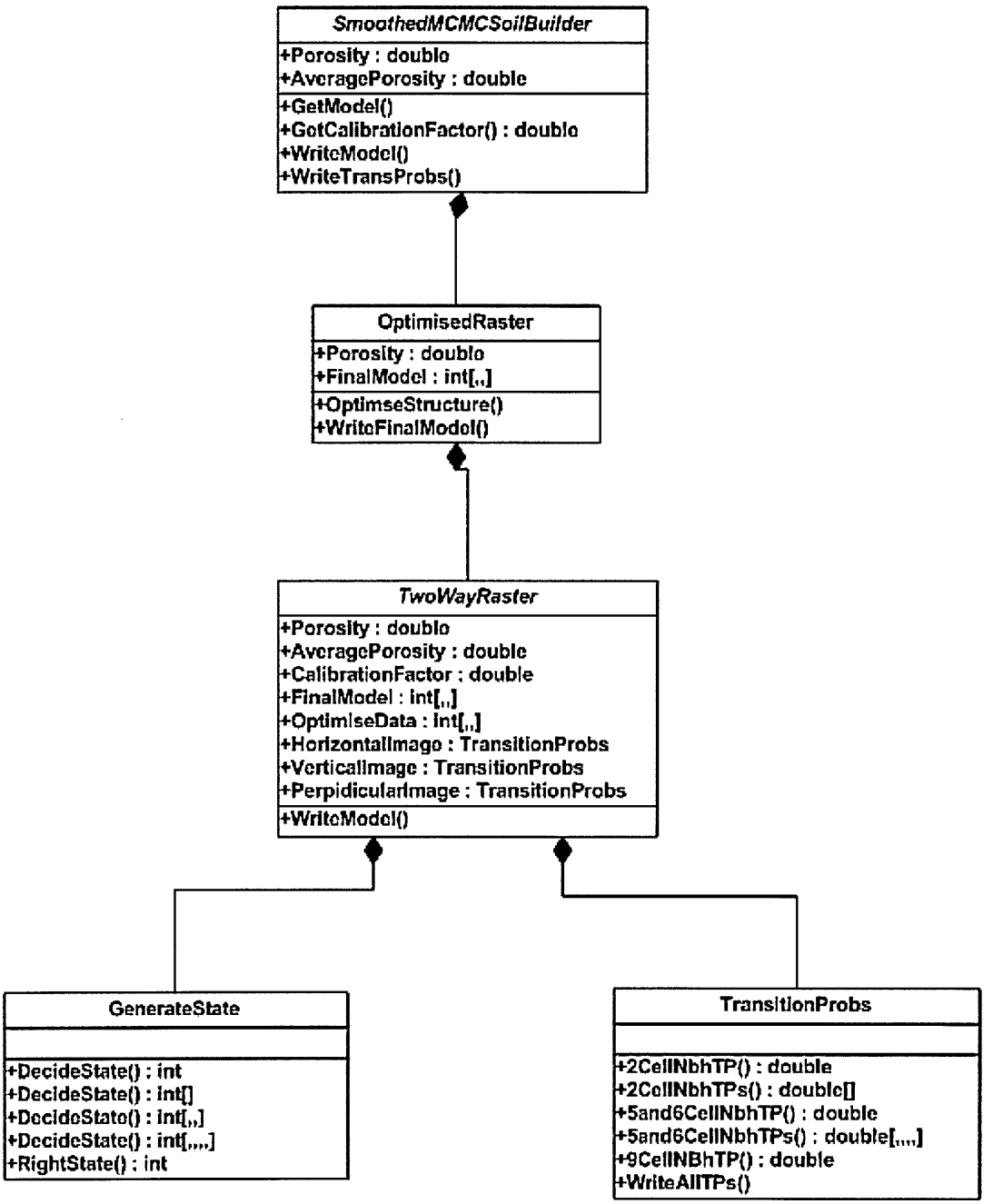
Wu, K., Van Dijke, I. M. J., Couples, G. D., Jiang, Z., Ma, J., Sorbie, K. S., Crawford, J. W., Young, I. and Zhang, X. 2006. 3D stochastic modelling of heterogeneous porous media - applications to reservoir rocks. *Transport in Porous Media,* 65: pp443 - 467

Young, I. M. and Crawford, J. W. 2001. Protozoan Life in a Fractal World. *Protist* 152: pp 123-126.

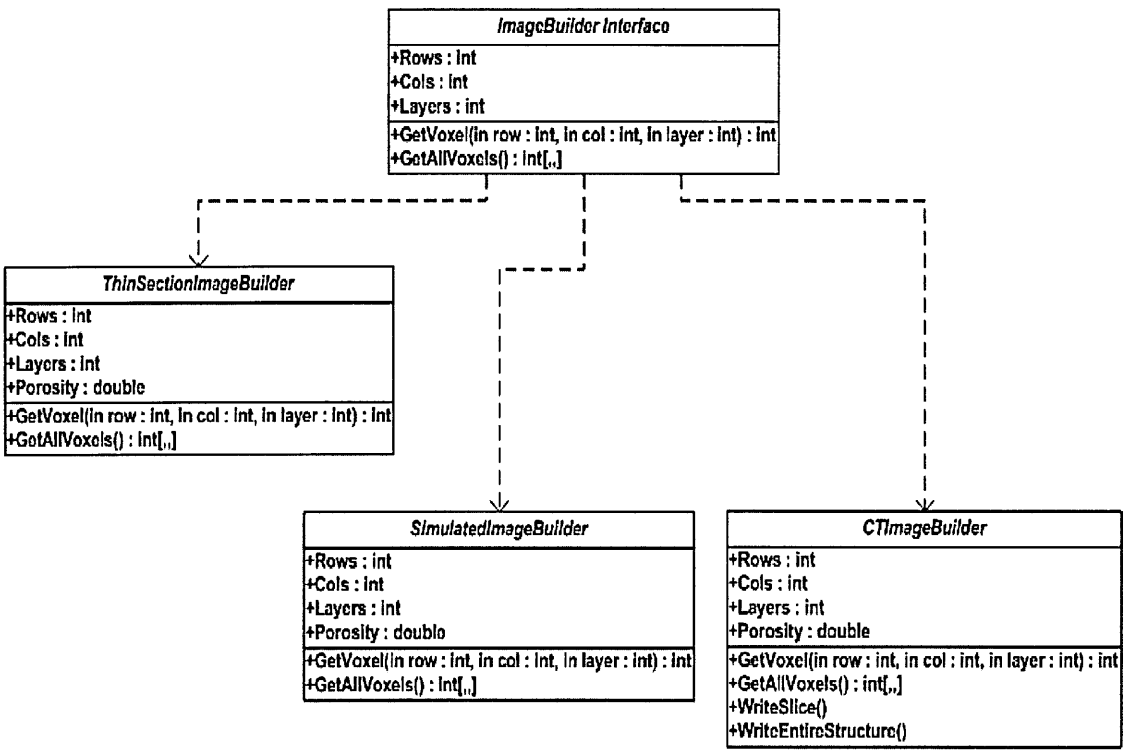
Young, I. M. and Crawford, J. W. 2004. Interactions and self-organization in the soil-micorbe complex. *Science.* 304: pp 1634 - 1637.

Appendix I – Class Designs

1.0 Soil Builder Object Model



2.0 Image Builder Object Model



Appendix II – Work Submitted for Publication

Simulation and Validation of Three-Dimensional Micro-Structure in Heterogeneous Media

J. M. Blair[1], R. E. Falconer[1], A. C. Milne[2], and I.M. Young[1], J. W. Crawford*[1]

[1] SIMBIOS Center, University of Abertay Dundee, DD1 1HG

[2] School of Computing and Creative Technologies, University of Abertay Dundee, DD1 1HG

*Corresponding author: **j.crawford@simbios.ac.uk**

ACKNOWLEDGMENTS

We gratefully acknowledge Dr Debbie Feeney, Dr Naoise Nunan, Prof Karl Ritz, Dr Mark Rivers, and Dr Kejian Wu for helpful discussions and technical support during this project. We also thank Biotechnology & Biological Sciences Research Council (BBSRC) for the provision of Strategic Studentship to JMB.

[redacted]

Media

ABSTRACT

Whilst it is now possible to image the 3D structure of soil using high resolution tomography, none of the techniques can simultaneously image the distribution of the resident soil microbes. This means that it is neither possible to visualise soil microbes in their natural habitat nor to model the consequences of soil structure for microbially-mediated processes. Biological thin sections offer the opportunity to simultaneously image microbes in structure but are necessarily restricted to 2D. Therefore a methodology is required to simulate 3D structures from 2D thin sections of soil that is extendable to simulate the spatial distribution of a range of soil components. Here, we present a model that is capable of using data gathered from 2D sections to predict the 3D structure of soil. An object-oriented approach to modeling is used to allow the individual representation of each structure voxel. This allows the model to encapsulate both data, presented here, and the subsequent addition of components such as microbial distribution and related diffusion-respiration processes, together in a 3D lattice of voxels. The model is validated using data derived from 3D x-ray tomography images of soil structure, and using 2D sections through that data set to predict 3D structure. A range of metrics are used to compare the modeled and imaged 3D structures. The comparison shows that the metrics for the modeled structures agree with those derived from the 3D images for higher porosities, but that systematic differences occur for the lowest porosity soils (< 11%). This is due to problems relating to the prediction of rare events such as the presence of large connected pores in low porosity samples.

Abbreviations: CT, Computed Tomography; 2D, two dimensional; 3D, three dimensional.

INTRODUCTION

All soils are arguably the most complex biomaterial on the planet. Soil is characterized by spatial heterogeneity on all scales that impact all nature and evolution of the physical and biological properties of the system (Crawford *et al.,* 2005). At the pore-scale, the physical heterogeneity promotes the coexistence of air and water, and by affecting the relative proportion of each is a major determinant of microbial activity and function (Crawford, 1994; Nunan *et al.,* 2002; Renault *et al.,* 2002; Vogel, Cousin & Roth, 2002). Ultimately, it is this microstructure that governs the flow of resources through the pore space of the system and creates a high diversity of microenvironments (Young & Crawford, 2004; Zhang *et al.,* 2004). Recently, it has been suggested that the role of microbial activity in the genesis of this structure and its impact on the resultant physical properties of soil results in a feedback between physical and biological processes that leads to self-organisation (Young & Crawford, 2004; Feeney *et al.,* 2006). Given the importance of the interaction between microbial and physical processes in soil, it is important that any conceptual framework investigating self-organisation should include a pore-scale structure model that can incorporate the distribution of microbes.

Recent progress in CT technology offers the possibility of imaging the 3D structure of soil, non-destructively and at resolutions of relevance to studying the interaction between physical and biological processes in soil (Nunan *et al.,* 2006). Unfortunately, current generation scanners and image processing software cannot resolve the microbial cells, and indeed the challenges presented preclude this for some time in the future. Currently, the only way to simultaneously image and quantify the distribution of microbes and soil structure is through the use of biological thin sections (Nunan *et al.,* 2003), but this is obviously restricted to 2D. Therefore in order to make progress, a methodology to extrapolate from 2D spatial data to 3D is required.

We have previously developed a methodology that simulates the 3D architecture of porous rocks based on images derived from 2D thin sections (Wu *et al.,* 2007). In that paper, the

method was tested by comparing the permeability, simulated mercury intrusion data and the topology of the pore space as determined by the Euler number (Vogel and Roth, 2001). The results for the rock samples were encouraging, but soil presents distinctive challenges for the methodology. Amongst the important differences with soil is the occurrence of coherent structures, e.g. pores, on a scale comparable to the sample, and the lack of a characteristic length-scale for the inherent heterogeneities therefore structural metrics depend on the scale of measurement (Bartoli *et al.*, 2005). Many probabilistic methods for modeling complex spatial structures, including our own, may fail when the structures have these properties simply because by definition large coherent structures are rare events and are notoriously difficult to predict when data is limiting (Kolar, 2004). For these reasons, the model of Wu *et al.*, (2007) must be separately validated for soil.

The metrics used for the validation must relate to the functional consequences of any differences between modeled and real structures. Only then can the significance of any difference be meaningfully interpreted. Where we are interested in the interaction between physical and biological processes in soil, the metrics should relate to how the model captures the features relevant to defining the microenvironment of soil microbes. These features include those that describe the transport of oxygen, the distribution of moisture and the available habitat for the microbes as well as the metrics associated with the scaling properties of the structure.

There are a number of metrics available for the characterisation of soil physical structure. Vogel and Krestzschmar (1996) digitally reconstructed the pore space from sequential thin sections of soil and skeletonized the results to derive the Euler-Poincare characteristic as an index of the 3-dimensional connectivity, of the pore space. Vogel *et al.*, (2002) used the same reconstruction technology to identify and described the pore geometry using pore-volume density, pore-surface density and the Euler number of the pore space. They showed, using a simulation of gas diffusion through the reconstructed pore space, that these metrics could be related to the effective diffusion coefficient of the structure. Using X-ray CT images of soil,

Perret *et al.,* (1999) characterized the pore network using metrics relating to tortuosity and connectivity, and showed that these could be related to convective fluid flow in soil.

The suites of metrics that have been applied to soil to date have not been designed specifically to characterize the interaction between physical and biological processes. Where microbes are imbedded in the soil matrix acting as sinks for diffusive and convective substrate, metrics that characterize the solid matrix and pore space simultaneously are required together with those that characterize diffusive and convective fluxes. In addition, we need to quantify the scaling properties of the structures in soil in recognition of the fact that sample size must be factored in to our interpretation of the images of structure.

The aim of this paper is to use CT-images of soil structure to validate a method for simulating the complex heterogeneity of the soil habitat as it impacts on the interaction between physical and biological processes. We propose the most relevant suite of metrics for the purpose. The model presented provides a theoretical framework to investigate the evolutionary ecology of the soil-microbe system.

METHODS

CT Imaged Soil Structures

We have used images collected from the study reported in Feeney *et al.,* (2006). In that work, a set of pot experiments was designed to investigate the potential for self-organization in the soil-microbe complex (Young & Crawford, 2004). Pots were filled with sandy loam soil (sand 59%, silt 19% and clay 7%, pH 5.6) and were either planted with a single Lolium perenne L. individual, or unplanted. The pots were incubated for 30 days and soil was destructively sampled at 8 time intervals during that period. Soil was sampled from the pots that did not contain plants (*control*), and from pots that contained plants at two locations: the *bulk* soil that did not contain roots; and *rhizosphere* soil close to the plant roots. Samples comprised aggregates of approximately 2mm in diameter and 5 of these aggregates from each of the

treatments at 0 and 30 days were imaged using the APS within the Argonne National Laboratory (Illinois, USA) in the GeoSoilEnviro Centre for Advanced Radiation Studies (GSECARS, station 13-BM-D). The resolution of the images was 4.4um and a thresholding algorithm was applied to convert the greyscale images into binary images with black corresponding to soil matrix and white corresponding to pore space. To avoid artefacts arising from the complex boundary shapes of the aggregates, a central portion of the images was extracted that corresponded to a cube of soil of side equal to 1.1mm. In the current study, we have used the CT images corresponding to control at zero days (denoted sample 1 hereafter), rhizosphere at zero days (sample 2), bulk soil at 30 days (sample 3) and rhizosphere soil at 30 days (sample 4). These samples were chosen because they span the full range of porosities observed in the experiment. Full details of the soils, experiments and imaging protocols can be found in Feeney *et al.,* (2006).

The Model

The method used to simulate the microstructure of soil is described fully in Wu *et al.,* (2007) and will be described again here. The method uses data in the form of 2D digitized images of soil sections to predict the structure of a 3D simulated volume of soil. The model has already been validated for a range of rock-types in Wu *et al.,* (2007) using morphological metrics. Here we test the applicability of the method to soil where we have both 2D and 3D images of the structure, and where we use metrics relevant to key soil processes.

The basis of the model is to use 2D data to estimate the probability that a particular voxel (3D pixel element) is in a particular state, conditioned on the states of voxels in a local 3D neighborhood. In principle, the state of a voxel could include whether it is pore or solid, and whether a microbe or cluster of microbes was present. Here, we focus on modelling structure alone, since the CT images do not capture microbes, and therefore the state of a voxel is defined to be either pore or solid. The method comprise a series of steps beginning with the derivation of the equivalent 2D conditional probabilities for each of the orthogonal sections. These probabilities are calculated by successively applying the six-cell neighborhood shown

in Figure 1a to each pair of 'target' pixels (i, j and i, j+1) of the image of the 2D section. For each combination of states of the target pixels, the number of instances of each of the observed combinations of states of the neighborhood cells is determined. The probability of realising a specific state of the target pixels, given the state of the neighborhood pixels, is estimated by dividing the number of occurrences of the given neighborhood state that coincide with the specific target cell state, and divide that by the total number in instances of the given neighborhood state. This process is repeated for each of the three orthogonal images.

The simulation of the 3D structure is initiated by first specifying the states of the cells in the bottom layer of the structure. This is done using the conditional probabilities derived for the horizontal section as described in Wu *et al.,* (2004). The states of the pixels along one edge of the layer are estimated by first choosing the state of a corner pixel randomly such that the probability that the pixel is a pore is given by the porosity. The state of the cell next to the corner cell is determined using the above method, applied to a two-cell neighborhood. Given the state of the corner cell, the state of the next cell is determined using the associated two-cell conditional probability. This is repeated until all pixels along the edge have been determined. The first pixel of the next row is determined using the two-cell conditional probability. The states of all other pixels in that row are determined using the six-cell neighborhood described above. This process is repeated until the states of all pixels in the first layer are determined.

Each further layer of the 3D lattice is obtained by combining the six-cell neighborhoods from each orthogonal section resulting in a localized neighborhood where the state of the target voxels depends on fifteen neighborhood states as shown in Figure 1b. Each orthogonal section offers an estimate of the state of one of the target voxels. Thus a means of combining these states to achieve the state of the voxel on the 3D lattice is required. Since the orthogonal slices are not spatially referenced there is no formal method for doing this. Wu *et al.,* (2007) has developed a weighting scheme for combining the states to achieve the voxel state on the 3D lattice. The important element of this weighting scheme is an optimization

step that determines the weighting factor by iteration until the porosity of the simulated 3D structure, ρ_{sim}, is sufficiently close to the average porosity of the 2D orthogonal sections, $\overline{\rho}_{2D}$, such that:

$$| \rho_{sim} - \overline{\rho}_{2D} | \leq 0.03 \qquad (1.0)$$

This algorithm is used to estimate the states of the voxels in all remaining layers, using modifications of the neighborhood to deal with voxels on the boundary in direct analogy with the description of the simulation of the 1st layer provided above, and detailed in Wu *et al.,* (2007). Finally, the 3D lattice of voxels is exposed to a smoothing algorithm that removes isolated voxels.

Functional Metrics Used to Quantify the Structure

To validate the model, comparisons between the simulated and CT-imaged structures are made using a number of metrics. These are porosity (ρ); the surface-connected porosity (ρ_{sc}); the percentage of surface-connected pore space ($\gamma = \rho_{sc} / \rho$); the frequency distribution of distances of matrix voxels to nearest surface-connected pore; the pore size distribution; and the fractal dimension. Each of these metrics has been chosen because it can be related to important physical and biological processes in soil.

Porosity is one of the most important metrics affecting physical transport rates. The porosity (ρ) of the simulated structures is important in identifying the reliability of the model. Although the model is calibrated using the average porosity of the orthogonal 2D sections, this does not guarantee that the porosity values corresponding to the simulated 3D structures will correspond to those of the CT-imaged structures.

The surface-connected porosity (ρ_{sc}) of a structure is the fraction of the soil volume comprising pore space that is connected to the boundary. This is a useful measure as it is the connected pore pathways that affect the transport of resources into the volume. The percentage of pore space that is surface-connected (γ) quantifies the partitioning of the pore space between isolated and boundary-connected pores. The computation of both these metrics requires that we define what constitutes a connection between adjacent voxels, and there are three ways this can be done. We denote the first of these neighborhoods as the 'faces' neighborhood, and this assumes that a voxel is connected to another voxel only if they meet at any face. The second connection neighborhood is denoted the 'edges' neighborhood, and this assumes that a voxel is connected to another voxel if it is in the faces neighborhood, or if they meet along any edge. The final connection neighborhood is denoted the 'vertices' neighborhood, and this assumes a voxel is connected to another if it is in its edges neighborhood, or if the voxels meet at any vertex. The sensitivity of the results to the choice of connection neighborhoods was investigated.

The distribution of solid voxels around surface-connected pores is a further metric used to quantify structures in this study. These distributions approximate the oxygen distributions in soil if each matrix voxel has the same potential respiration rate (Rappoldt & Crawford, 2001). The frequency distribution of voxel distances to the nearest connected pore is calculated for each structure sample, both CT and simulated, allowing comparisons between the two sets of data to be made.

The final metric we consider is the fractal dimension of the pore space, which is a measure of its clustering properties (Pachepsky, Crawford & Rawls, 2000). The method employed here uses the box counting method to estimate the fractal (strictly the Hausdorff) dimension (Baveye, Parlange & Stewart, 1998). A log-log plot of the number of boxes required to cover the pattern against box size gives a straight line whose slope is the fractal dimension. Whilst the fractal dimension gives an estimate of the clustering of the space, it does not give a measure of the density of the pores. Therefore, structures with many small pores could have the same fractal dimension as structures with a few larger pores. The intercept of the plotted

line described above with the vertical axis gives a measure (called the lacunarity) of the density of the pore space, and a pattern with the same fractal dimension but with smaller pores would have a larger intercept (Mandelbrot, 1977). Therefore, both the slope and the height of the plot characterize the distribution.

RESULTS

Porosity (ρ)

This porosity data (Table 1) shows that the modeled structures capture the variation of porosity found across the CT-imaged sets. Statistical t tests were carried out on the porosity results and three of the four simulated structures are not significantly different from the CT structures ($p > 0.05$). However, although not significant there is a general trend to underestimate the porosity of the simulated structure compared to the CT-imaged data sets. Further sample 1 (lowest porosity structure) has a p value of 0.026 ($p < 0.05$) therefore we reject the Null Hypothesis that there is no difference between the datasets and conclude that the low porosity simulated structure is significantly different from the CT-imaged sample. Figure 5 shows a 3D visualization of the CT-imaged and simulated structures for sample 1.

Surface-connected Porosity (ρ_{sc})

For three of the four samples for each of the faces, edges and vertices neighborhood schemes, used to calculate the surface connected porosity, the t test results indicate no significant difference between the simulated and CT samples ($p > 0.05$). However statistical analysis of sample 3 infers that the simulated and CT samples are significantly different for the faces and edges neighborhoods ($p = 0.03$, $p = 0.01$ respectively) but not for the vertices neighborhood ($p = 0.402$). Further the ρ_{sc} metric calculated for the lower porosity simulated structures (sample sets 1 and 2) vary more from the values for the corresponding CT-imaged structures.

Percentage Of Pore Space Surface-connected (γ)

These results are consistent with the surface connected porosity results. For three of the four samples for each of the faces, edges and vertices neighborhood schemes, used to calculate the percentage of pore space surface connected, the t test results indicate no significant difference between the simulated and CT samples ($p > 0.05$). However statistical analysis of sample 3 infers that the simulated and CT samples are significantly different for the faces neighborhood ($p = 0.03$) but not for the edges and vertices neighborhoods ($p = 0.06$, $p = 0.15$ respectively). Again there is a general trend for the simulated percentage of pore space surface-connected to be underestimated. As expected, the connection neighborhoods that offer each voxel a larger number of possible connected neighbours results in higher values for the ρ_{sc} and γ metrics.

Frequency Distributions

For each CT-imaged sample the frequency distribution of the distance to nearest connected pore for each voxel was calculated. The frequency distributions obtained from the simulated structures follow similar qualitative trends to the CT-imaged structures. It was found that high porosity structures (sample sets 3 and 4) resulted in maximum distances to nearest connected pore of approximately half the maximum distances found in lower porosity structures (sample sets 1 and 2). It can be seen that there are a greater number of voxels with distances of less than 135 microns to a connected pore in the high porosity samples than in the low porosity samples (Figure 2). This is a consequence of the higher surface-connected porosities of these structures.

The agreement between the distributions for the simulated and CT structures is better for short distances than it is for larger distances, this is evident from Figure 2 as the simulated and CT derived distributions overlap for small distance but diverge for larger ones. There is no general trend towards over or underestimation as porosity increases. Larger distance

measures will be sensitive to the simulated distribution of smaller pores and there is evidence from the next set of results that the simulation is weakest in the prediction of the small scale pore space.

Fractal Dimension and Pore Size Distributions

Two further measures of the soil structure were calculated to aid in analysis of the structures, these were the fractal dimension and pore size distribution. Both these measures give an indication of the clustering properties of the pore space. The fractal dimensions obtained for the simulated and CT-imaged structures are not significantly different ($p > 0.05$) as can be seen from Table 2. Figure 4 illustrates the results of the box counting method for CT-imaged and simulated structures with a low and high porosity. The slope of the lines characterising the fractal dimension for CT-imaged and simulated sample 1 are 2.69 and 2.84 respectively. The fractal dimension for the CT-imaged and simulated sample 4 are 2.79 and 2.80 respectively.

The pore size distribution for the simulated and CT-imaged structures is shown in Figure 3. The distribution for the simulated structure for the higher porosity case (sample 4) is closer to that of the corresponding CT-imaged structure, than in the lower porosity case (sample 1).

DISCUSSION

The results presented show that the model produces simulated structures that capture the variation in structure found across the real CT-imaged structures. The analysis carried out on the CT-imaged data samples and the simulated structures show a distinct split between high porosity (> 11%) and low porosity (< 11%) treatments.

The simulations performed well for the high porosity samples (> 11 %), indeed for these samples there is excellent agreement between the simulated and CT structures as quantified by the metrics described in this paper. There is however a tendency for the ρ, ρ_{sc} and γ metrics to be underestimated although this was not significant.

For extremely low porosity vales $\rho \leq 6\%$ the structures were not so good at replicating the bulk porosity. Further, unless the vertices neighborhood was used in the calculation of ρ_{sc} and γ metrics the simulated structures were not similar to the CT structures for $\rho \leq 8.1$ %. Furthermore for $\rho \leq 8.1$ the simulated pore size distributions varied greatly from the CT pore size distributions. The most likely reason for these discrepancies is the inability of the method to simulate rare events, particularly the appearance of large-scale correlated structure in low porosity media.

Our analysis showed that in general, the higher porosity structures had more surface-connected pathways than lower porosity sets (Table 1). Furthermore the high porosity sets show greater numbers of soil voxels concentrated around surface-connected pores than in lower porosity structures, with the maximum distance to a connected pore in high porosity structure sets being approximately half the maximum distance found in lower porosity sets (Fig. 2). As this distribution relates to the distribution of oxygen in the soil matrix, the correspondence between the distributions from the simulated and CT structures, particularly for smaller distances, suggests that the model should be able to capture the functionally important characteristics of high porosity structures.

The use of a detection neighborhood for computing the various porosity measures was investigated and, as expected, the larger neighborhoods resulted in an increase in the ρ_{sc} and γ measures resulting in convergence of agreement between the simulated and CT structures. This outcome is due to the larger neighborhoods providing more possible connections between voxels therefore compensating for the general underestimation of the metrics. It can be noted from Table 1 that the different detection neighborhoods used for computing the ρ_{sc} metric on CT-imaged and simulated structures showed little change between the faces and edges algorithm and the largest effect was between the edges and vertices algorithm for both simulated and CT structures.

The fractal dimension measurements show good agreement between CT and simulated structures for all porosities. The gradients and heights of the lines of the log-log plots show close agreement between CT-imaged and simulated structures therefore the model can replicate the clustering properties of the CT-imaged structure. As the porosity increases so does the fractal dimension. Overall, there is a general trend across all metrics where higher porosity simulations show better correlation with the CT-imaged structures than lower porosity sets.

CONCLUSIONS

The work presented here has used CT-imaged soil structures to validate a method of simulating the 3D soil habitat. It has been shown that the model adequately captures the pore architecture of structures where the porosity is above 11%. Below this value the simulated structures capture the main porosity values but not the detailed characteristics of porosity such as connectivity and clustering properties as characterised by the fractal dimension. The reason for this failure is due to the inability of the model to simulate the rare events such as large pores in low porosity structures. The fact that we have focussed here on small samples at high resolution means that the porosities are substantially smaller than normally found in soil at larger scales. Our results suggest that the model would work well at these scales and porosities.

Nevertheless, the model has been found to simulate the general trends of the soil microstructure found across different porosity classes. The simulated structures provide the 3-dimesional-habitat space in which a theoretical framework to study the evolutionary ecology of the soil-microbe system can be built.

REFERENCES

Baveye, P., J. Parlange, and B.A. Stewart. 1998. *Fractals in soil science*. CRC Press.

Bartoli, F., Genevois-Gomendy,V., Royer, F.F, Niquet, S., Vivier, H and Grayson, R. 2005. A multiscale study of silty soil structure. *European Journal of Soil Science.* 56 (2), 207 -224.

Crawford, J.W., J.A. Harris, K. Ritz, and I.M. Young. 2005. Towards an evolutionary ecology of life in soil. *TRENDS in Ecology and Evolution*. 20(2):81-87.

Crawford, J.W. 1994. The relationship between structure and the hydraulic conductivity of soil. *European Journal of Soil Science*. 45:493-502.

Denef, K. and J. Six. 2005. Clay mineralogy determines the importance of biological versus abiotic processes for macroaggregate formation and stabilization. *European Journal of Soil Science*. 56(4): 469-479.

Feeney, D. S., J.W. Crawford, T. Daniall, P.D. Hallett, N. Nunan, K. Ritz and I.M. Young. 2006. 3d micro-organisation of the soil-root-microbe system. *Microbial Ecology (in press)*.

Kolar, C. 2004. Risk assessment and screening for potentially invasive fishes. *New Zealand Journal of Marine and Freshwater Research*. Vol. 38: 391–397

Mandelbrot, B.B., 1977. The fractal geometry of nature. *W. H. Freeman and Company, New York.*

Nunan, N., K. Ritz, M. Rivers, D.S. Feeney and I.M. Young. 2006. Investigating microbial micro-habitat structure using x-ray computed tomography. *Geoderma (in press)*.

Nunan, N., K. Wu,I.M. Young, J.W. Crawford, and K. Ritz. 2003. Spatial distribution of bacterial communities and their relationship with the micro-architecture of soil. *FEMS Microbiology Ecology*. 44:203-215.

Nunan, N., K. Wu, I.M. Young, J.W. Crawford and K. Ritz. 2002. In situ spatial patterns of soil bacterial populations, mapped at multiple scales, in an arable soil. *Microbial Ecology*. 44:296-305.

Pachepsky, Y. A., J.W. Crawford, and W.J. Rawls. 2000. *Fractals in soil science*. Elsevier.

Perret, J., S.O. Prasher, A. Kantzas, and C. Langford. 1999. Three-dimensional quantification of macropore networks in undistrurbed soil cores. *Soil Science Society of America*. 63:1530-1543.

Rappoldt, C. and J.W. Crawford. 2001. An approximation for the anoxic fraction of fractal soil. In: Reynolds, W. D., Drury, C. F. and Tan, C. S. eds. 4th Eastern Canada Soil Structure Workshop, Greenhouse and Processing Crops Research Centre, Harrow, Ontario, Canada.

Renault, P., K. Khalil, F. Dassonville, D. Mohrath, R. Lensi, F. Chadoeuf, C. Chenu, L. Bidel and F. Lafolie. 2002. Interactions between gas transport and biochemical processes: The effect of the soil structure. 17th WCSS, Thialand, August 2002. p.922.

Vogel, H. J., I. Cousin, and K. Roth. 2002. Quantification of pore structure and gas diffusion as a function of scale. *European Journal of Soil Science*. 53(3):465-473.

Vogel, H. J. and Roth, K.: 2001, Quantitative morphology and network representation of soil pore structure, *Adv. Water Resour.* **24**, 233–242.

Vogel, H. J. and A. Kretzschmar. 1996. Topological characterization of pore space in soil - sample preparation and digital image-processing. *Geoderma*. 79:23-38.

Wu, K., N. Nunan, K. Ritz, J.W. Crawford, J.M. Milne and I.M. Young. 2007. Quantitative characterisation of 3-d heterogeneous porous media using mutli-markov chain model: A case study with soil. *Soil Science Society of America (in press).*

Wu, K., Nunan, N., Ritz, K., Young, I., Crawford, J. 2004. An efficient Markov chain model for the simulation of heterogeneous soil structure. *Soil Sci. Soc. Am. J.* 68: 346-351

Young, I. M. and J.W. Crawford. 2004. Interactions and self-organization in the soil-micorbe complex. *Science.* 304(5677):1634-1637.

Zhang, X., L.K. Leeks, A.G. Bengough, J.W. Crawford and I.M Young. 2005. Determination of soil hydraulic conductivity with the lattice boltzmann method and soil thin-section technique. *Journal of Hydrology.* 306:59-70.

Figure 1 – Neighborhoods for simulating structure. (a) Six-cell neighborhood, (b) Fifteen-cell neighborhood. Gray scale illustrate target cells used in algorithm.

Figure 2 – Frequency distributions of voxel distances to nearest connected pore for sample 1 and 4 (S1 and S4 respectively) for CT-imaged data (CT) and simulated data (Sim).

Figure. 3 – Example of pore size distributions for sample 1 and 4 (S1 and S4 respectively) for CT-imaged data (CT) and simulated data (Sim).

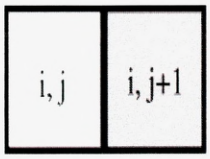
Figure 4 – Example of fractal dimension for sample 1 and 4 (S1 and S4 respectively) for CT-imaged data (CT) and simulated data (Sim)

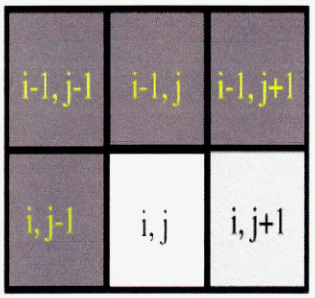
Figure 5 – Comparison of CT-imaged (a) and simulated structure (b) for one of the replicates of Sample 1.

Table 1. Averaged porosity metrics obtained for CT-imaged and simulated structures. Standard error shown in brackets.

Table 2 Comparison of averaged fractal dimension for CT-imaged and simulated structures. Standard error in brackets.

Fig.1a.


i, j
i, j+1


i-1, j-1
i-1, j
i-1, j+1
i, j-1
i, j
i, j+1

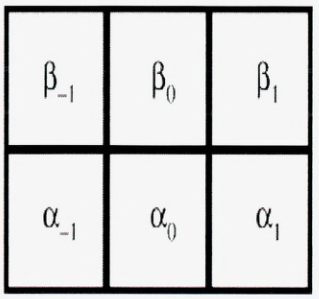

β_{-1}
β_0
β_1
α_{-1}
α_0
α_1

Fig. 1b.

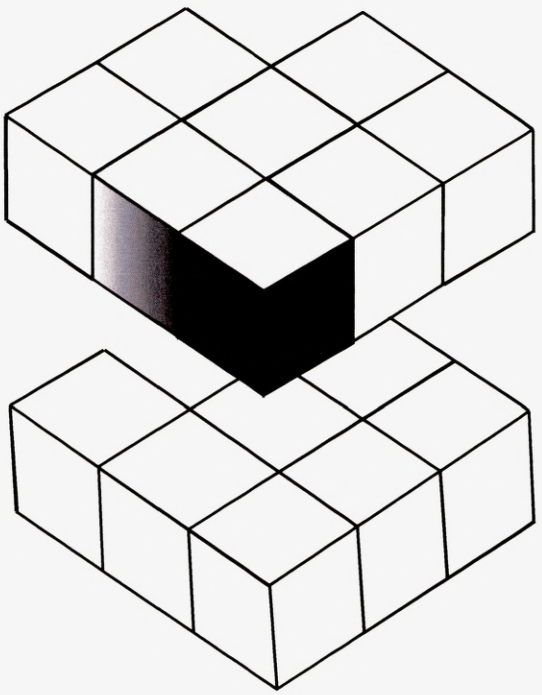


Fig. 2

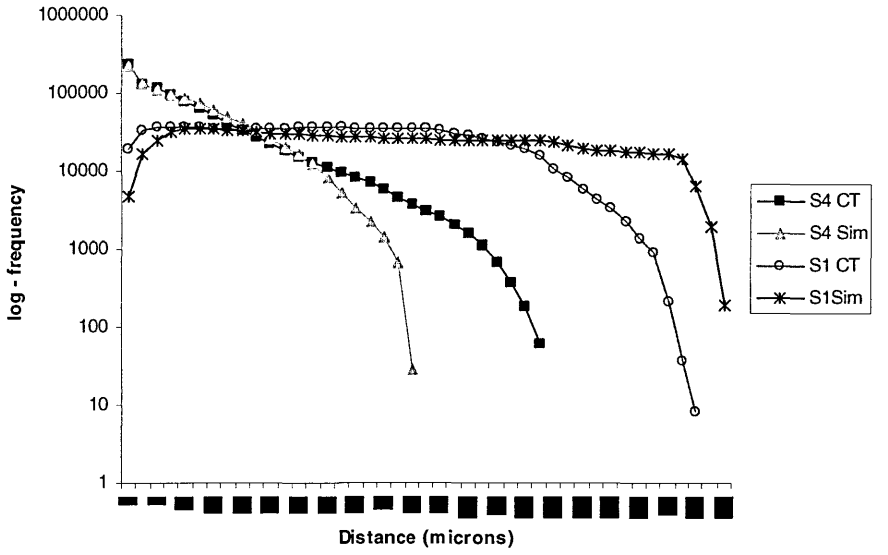

1000000
100000
10000
1000
100
10
1
log - frequency
Distance (microns)
S4 CT
S4 Sim
S1 CT
S1Sim

Fig. 3

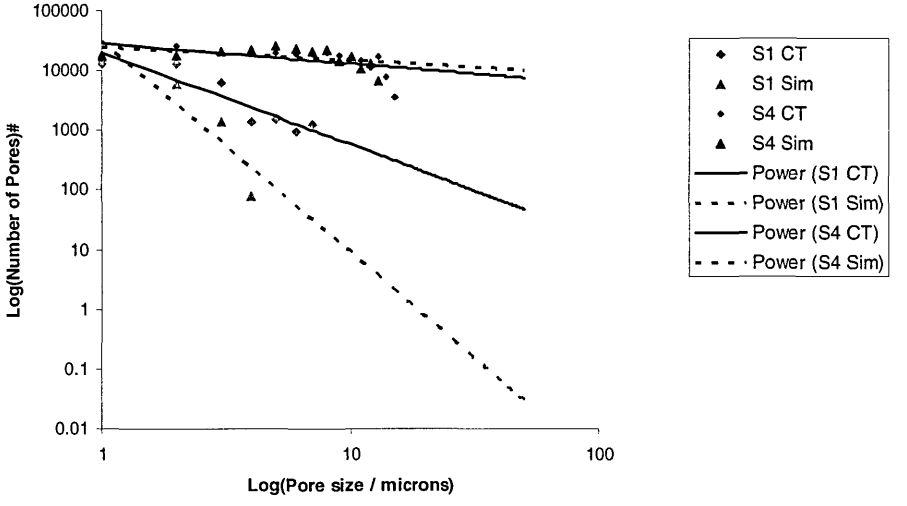

100000
10000
1000
100
10
1
0.1
0.01
Log(Number of Pores)#
1
10
100
Log(Pore size / microns)
S1 CT
S1 Sim
S4 CT
S4 Sim
Power (S1 CT)
Power (S1 Sim)
Power (S4 CT)
Power (S4 Sim)

Fig. 4

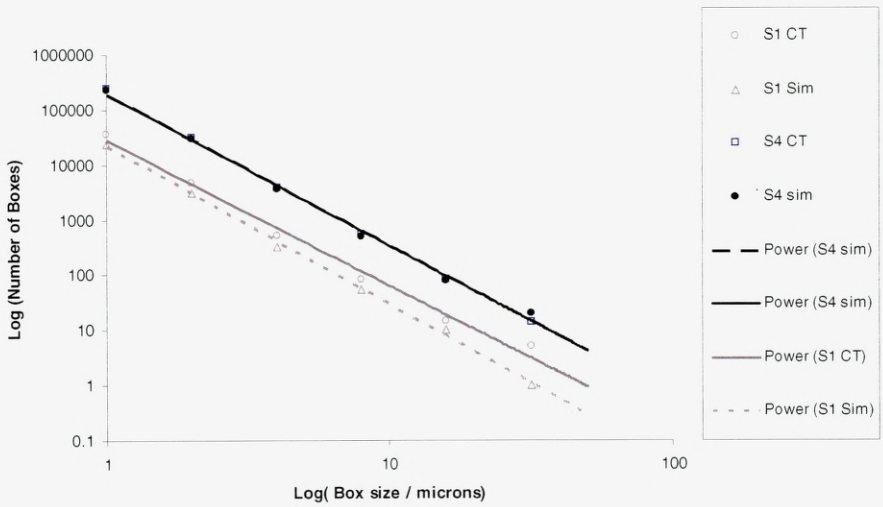
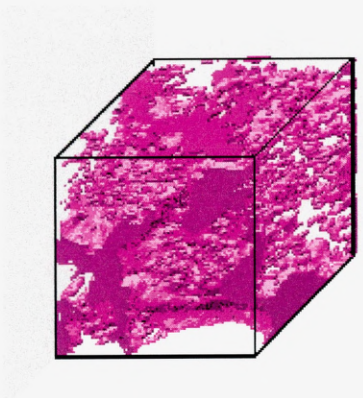
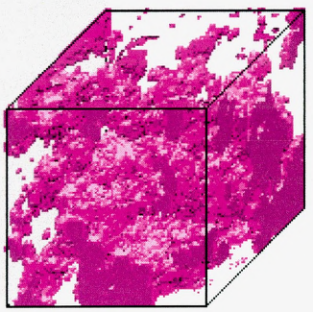

1000000
100000
10000
1000
100
10
1
0.1
1
10
100
Log (Number of Boxes)
Log(Box size / microns)
S1 CT
S1 Sim
S4 CT
S4 sim
Power (S4 sim)
Power (S4 sim)
Power (S1 CT)
Power (S1 Sim)

Fig 5



(a)



(b)

Table 1. Averaged porosity metrics obtained for CT-imaged and simulated structures. Standard error shown in brackets.

	Porosity		Surface-connected Porosity						% Pore Space Surface-connected					
	CT	Simulated	CT			Simulated			CT			Simulated		
Detection Neighborhood			Faces	Edges	Vertices	Faces	Edges	Vertices	Faces	Edges	Vertices	Faces	Edges	Vertices
Sample 1	5.1 (0.6)	4.4 (0.9)	1.6 (0.6)	1.6 (0.6)	2.9 (1.0)	0.7 (0.2)	0.8 (0.3)	1.4 (0.5)	28.6 (8.6)	28.9 (8.5)	50.1 (12.4)	15.5 (3.2)	16.5 (3.2)	26.1 (6.0)
Sample 2	5.7 (1.0)	5.4 (1.7)	3.7 (0.9)	3.7 (0.9)	4.0 (1.0)	2.9 (1.9)	3.0 (2.0)	3.1 (2.0)	64.3 (4.0)	64.4 (4.0)	67.4 (5.7)	36.5 (18.8)	36.9 (18.9)	39.9 (18.5)
Sample 3	8.1 (2.4)	7.9 (1.7)	5.4 (1.9)	5.4 (1.9)	5.8 (2.0)	3.5 (1.9)	3.9 (2.1)	5.2 (2.10)	60.6 (10.0)	60.7 (10.0)	64.9 (9.1)	33.63 (13.0)	38.33 (14.3)	53.0 (14.2)
Sample 4	15.1 (5.4)	14.6 (3.7)	12.8 (4.1)	12.9 (4.1)	13.1 (4.0)	11.7 (4.5)	11.8 (4.5)	12.9 (4.1)	71.6 (13.2)	72.0 (13.2)	74.1 (12.2)	61.0 (16.2)	63.5 (15.5)	76.7 (10.8)

Table 2 Comparison of averaged fractal dimension for CT-imaged and simulated structures. Standard error in brackets.

	Fractal Dimension	
	CT	Simulated
Sample 1	2.75 (0.03)	2.79 (0.02)
Sample 2	2.93 (0.04)	2.88 (0.04)
Sample 3	2.84 (0.05)	2.89 (0.05)
Sample 4	2.86 (0.56)	2.86 (0.02)